A POSTERIORI ERROR ESTIMATES FOR DISCONTINUOUS GALERKIN METHODS USING NON-POLYNOMIAL BASIS FUNCTIONS. PART I: SECOND ORDER LINEAR PDE

Lin Lin[1] and Benjamin Stamm[2]

Abstract. We present the first systematic work for deriving a posteriori error estimates for general non-polynomial basis functions in an interior penalty discontinuous Galerkin (DG) formulation for solving second order linear PDEs. Our residual type upper and lower bound error estimates measure the error in the energy norm. The main merit of our method is that the method is parameter-free, in the sense that all but one solution-dependent constants appearing in the upper and lower bound estimates are explicitly computable by solving local eigenvalue problems, and the only non-computable constant can be reasonably approximated by a computable one without affecting the overall effectiveness of the estimates in practice. As a side product of our formulation, the penalty parameter in the interior penalty formulation can be automatically determined as well. We develop an efficient numerical procedure to compute the error estimators. Numerical results for a variety of problems in 1D and 2D demonstrate that both the upper bound and lower bound are effective.

.

1. Introduction

Let Ω be a bounded domain. We consider the development of a posteriori error estimates for the following second order linear PDE

$$-\Delta u + Vu = f, \qquad \text{in } \Omega, \tag{1}$$

using the discontinuous Galerkin (DG) formulation with general non-polynomial basis sets.

Such equation arises in many scientific and engineering problems such as in electromagnetism, geophysics, quantum physics, to name a few. In order to solve Eq. (1) in practice, it is desirable to reduce the number of degrees of freedom for discretizing Eq. (1) to have a smaller algebraic problem to solve. While standard polynomial basis functions can approach a complete basis set and is versatile enough to represent almost any function of interest, the resulting number of degrees of freedom is usually large even when high order polynomials are used. Non-polynomial basis functions are therefore often employed to reduce the number of degrees of freedom, and are widely used to solve Eq. (1) and other equations, including the planewave basis set for solving Helmholtz equation [12, 26], the heterogeneous

[1] Department of Mathematics, University of California Berkeley and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: linlin@math.berkeley.edu

[2] Sorbonne Universités, UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005 Paris, France. Email: stamm@ann.jussieu.fr

multiscale method (HMM) [7] and the multiscale finite element method [13] for solving multiscale elliptic equations, and the various non-polynomial basis set used in quantum chemistry such as the Gaussian basis set [10], atomic orbital basis set [16], and adaptive local basis set [20] etc.

Besides solving the equation, it is also often desirable to assess the accuracy of the numerical solution via a posteriori error estimates and to design approximation spaces that result in a uniform distribution of the error in space to achieve best accuracy for a given number of degrees of freedom. In this paper we focus on the a posteriori error estimates in the interior penalty DG formulation [3–5, 8, 21, 27].

The DG formulation has the advantage that it formally relaxes the continuity constraint of basis functions at the inter-element boundary, and is therefore particularly suitable for incorporating general basis functions, which are difficult to match at the inter-element boundary.

1.1. **Previous work**

Compared to the many existing works on a posteriori error estimates using polynomial basis functions in the DG formulation [14, 17, 23], it is much more difficult to develop systematic a posteriori error estimates for general non-polynomial basis functions. One of the important reasons is that approximation and scaling properties of the function space spanned by non-polynomial basis functions, which are key to a posteriori error estimates, are generally difficult to deduce. For instance, Amara et al [2] developed the upper bound error estimates for the Helmholtz equation in planewave basis enriched DG method, and the error is measured in the L^2–norm. Kaye et al [18] developed the upper bound error estimates for solving linear eigenvalue problems using non-polynomial basis functions in a DG framework, which generalizes the work of Giani et al [11] for polynomial basis functions. However, the assumption of approximation properties on the function space is in general difficult to verify. Though not in the DG framework, Ohlberger [22] developed the a posteriori error estimates for the HMM method for elliptic homogenization problems.

The difficulties of a posteriori error analysis for general non-polynomial basis functions are largely due to the lack of credible methods for measuring the ratio of the error using different norms, defined in proper function spaces. For instance, approximately speaking, in a residual based error estimator, the constants associated with the residual requires the estimation of ratio of the error measured using L^2–norm and the H^1–norm. The scaling properties of such constants with respect to the increase of the number of basis functions on a particular element can be rather intrigue for non-polynomial basis functions. The estimation of such constants is already complicated for polynomial basis functions or planewave basis functions, not to mention the case when the non-polynomial basis functions come from numerical solution without a analytic recipe, or even worse, the basis functions do not in practice form a complete basis set with only saturating accuracy.

1.2. **Contribution**

To the extent of our knowledge, this is the first systematic work for deriving a posteriori error estimates for general non-polynomial basis functions in a DG framework. Our upper and lower bound error estimates are residual type estimators for the error in the energy norm. In our formulation, all but one basis-dependent constants appearing in the upper and lower bound estimates are explicitly computable by solving local eigenvalue problems. For solution with sufficient regularity (for instance $u \in H^2(\Omega)$), the only non-computable constant can be reasonably approximated by a computable one without affecting the overall effectiveness of the estimates. While the requirement of $H^2(\Omega)$ regularity appears to be a formal drawback in the context of a posteriori error estimates, the main goal of this work is to develop a posteriori error estimates for general basis sets rather than for h-refinement, and the difficulty of general basis sets holds even if the solution has $C^\infty(\Omega)$ regularity. Therefore we think our method can have important practical values. As a side product of our results, the penalty parameter in the interior penalty formulation is also automatically computed, and the computed constants guarantees that the coercivity of the resulting DG bilinear form for Poisson's equation.

We develop an efficient numerical procedure to compute these constants. Both the formulation and the practical implementation of our method are independent of how the basis functions are generated. Although the numerical procedure is developed for general non-polynomial basis functions, we find that the procedure, when applied to standard polynomial basis functions, generates constants are even more accurate than the analytical asymptotic result. Numerical results for a variety of problems in 1D and 2D

indicate that both the upper bound and lower bound are sharp, and the effectiveness of the estimators holds even at the level of each element.

1.3. Outline

The rest of the paper is organized as follows. After an introduction to some technical results in section 2, we start with the derivation of the upper bound a posteriori error estimates for the Poisson's equation in section 3, without the potential term V. We then generalize the derivation of the upper bound error estimates to indefinite problems with the potential term, as well as the lower bound error estimates in section 4. We elaborate in section 5 on the numerical methods for computing the constants appearing in the upper and lower bound estimates needed in our analysis. Finally, we present numerical results in section 6, before we conclude in section 7 followed by an appendix.

2. Preliminary results

2.1. Mesh, broken spaces, jump and average operators

Let $\Omega = (0,1)^d$, $d = 1, 2, 3$ and let $\mathcal{K}$ be a regular partition of Ω into elements $\kappa \in \mathcal{K}$. That is, we assume that the interior of $\overline{\kappa} \cap \overline{\kappa}'$, for any $\kappa, \kappa' \in \mathcal{K}$, is either an element of $\mathcal{K}$, a common face, edge, vertex of the partition or the empty set. For simplicity, we identify the boundary of Ω in a periodical manner. That means, that we also assume the partition to be regular across the boundary $\partial\Omega$. We remark that although the assumption of a rectangular domain with periodic boundary condition appears to be restrictive, such setup already directly finds its application in important areas such as quantum chemistry and materials science. However, the analysis below is not restricted to equations with periodic boundary condition. Other boundary conditions, such as Dirichlet or Neumann boundary conditions can be employed as well with minor modification. Generalization to non-rectangular domain does not introduce conceptual difficulties either, but may lead to changes in numerical schemes for estimating relevant constants in section 5, if the tensorial structure of the grid points is not preserved.

Let $N = (N_\kappa)_{\kappa \in \mathcal{K}}$ denote the vector of the local number of degrees of freedom N_κ on each element $\kappa \in \mathcal{K}$. Let $\mathbb{V}_N = \bigoplus_{\kappa \in \mathcal{K}} \mathbb{V}_N(\kappa)$ by any piecewise discontinuous approximation space on a partition $\mathcal{K}$ of the domain Ω. It is important to highlight that little is known about the a priori information of $\mathbb{V}_N$ except that we assume that each $\mathbb{V}_N(\kappa)$ contains constant functions and that $\mathbb{V}_N(\kappa) \subset H^{\frac{3}{2}}(\kappa)$, so that the traces of ∇v_N on the boundary $\partial\kappa$ are well-defined for all $v_N \in \mathbb{V}_N(\kappa)$, for all $\kappa \in \mathcal{K}$. We denote by $H^s(\kappa)$ the standard Sobolev space of $L^2(\kappa)$-functions such that all partial derivatives of order $s \in \mathbb{N}$ or less lie as well in $L^2(\kappa)$. By $H^s(\mathcal{K})$, we denote the set of piecewise H^s-functions defined by

$$H^s(\mathcal{K}) = \left\{ v \in L^2(\Omega) \,\middle|\, v|_\kappa \in H^s(\kappa), \forall \kappa \in \mathcal{K} \right\},$$

also referred to as the broken Sobolev space. We denote by $H^1_\#(\Omega)$ the space of periodic H^1-functions on Ω. We further define the element-wise resp. face-wise scalar-products and norms as

$$(v, w)_\mathcal{K} = \sum_{\kappa \in \mathcal{K}} (v, w)_\kappa \qquad \text{and} \qquad \|v\|_\mathcal{K} = (v, v)_\mathcal{K}^{\frac{1}{2}}.$$

The L^2-norm on κ and Ω are denoted by $\|\cdot\|_\kappa$ and $\|\cdot\|_\Omega$, respectively.

The jump and average operators on a face $\overline{F} = \overline{\kappa} \cap \overline{\kappa}'$ are defined in a standard manner by

$$\{\!\{v\}\!\} = \tfrac{1}{2}(v|_\kappa + v|_{\kappa'}), \qquad\qquad \text{and} \qquad [\![v]\!] = v|_\kappa n_\kappa + v|_{\kappa'} n_{\kappa'},$$
$$\{\!\{\nabla v\}\!\} = \tfrac{1}{2}(\nabla v|_\kappa + \nabla v|_{\kappa'}), \qquad \text{and} \qquad [\![\nabla v]\!] = \nabla v|_\kappa n_\kappa + \nabla v|_{\kappa'} n_{\kappa'},$$

where n_κ denotes the exterior unit normal of the element κ. Further we state some standard results

Finally we recall the standard result of piecewise integration by parts formula that will be employed several times in the upcoming analysis.

Lemma 2.1. *Let* $v, w \in H^2(\mathcal{K})$. *Then, there holds*

$$\sum_{\kappa \in \mathcal{K}} \left[(\Delta v, w)_\kappa + (\nabla v, \nabla w)_\kappa \right] = \tfrac{1}{2} \sum_{\kappa \in \mathcal{K}} \left[([\![\nabla v]\!], w)_{\partial\kappa} + (\nabla v, [\![w]\!])_{\partial\kappa} \right].$$

2.2. Projections

For any element $\kappa \in \mathcal{K}$, let us denote by $\Pi_0^\kappa : L^2(\kappa) \to \mathbb{R}$ the $L^2(\kappa)$-projection onto constant functions defined by

$$(\Pi_0^\kappa v, w)_\kappa = (v, w)_\kappa, \qquad \forall w \in \mathbb{R},$$

that is explicitly given by $\Pi_0^\kappa v = \frac{1}{|\kappa|} \int_\kappa v \, dx$. On $H^1(\kappa)$ we define the following scalar product and norm

$$(v, w)_{\star,\kappa} = (\Pi_0^\kappa v, \Pi_0^\kappa w)_\kappa + (\nabla v, \nabla w)_\kappa, \tag{2}$$

$$\|v\|_{\star,\kappa} = (v, v)_{\star,\kappa}^{\frac{1}{2}},$$

for all $v, w \in H^1(\kappa)$ and the corresponding projection $\Pi_N^\kappa : H^1(\kappa) \to \mathbb{V}_N(\kappa)$ by

$$(\Pi_N^\kappa v, w_N)_{\star,\kappa} = (v, w_N)_{\star,\kappa} \qquad \forall w_N \in \mathbb{V}_N(\kappa). \tag{3}$$

Then, it is easy to see that this projection satisfies the following properties

$$(v - \Pi_N^\kappa v, c)_\kappa = 0, \qquad\qquad \forall c \in \mathbb{R}, \forall v \in H^1(\kappa),$$

or equivalently expressed as $\Pi_0^\kappa(v - \Pi_N^\kappa v) = 0$. This implies that

$$(\nabla(v - \Pi_N^\kappa v), \nabla w_N)_\kappa = 0, \qquad \forall w_N \in \mathbb{V}_N(\kappa), \forall v \in H^1(\kappa), \tag{4}$$

$$\|\nabla(v - \Pi_N^\kappa v)\|_\kappa \le \|\nabla v\|_\kappa, \qquad \forall v \in H^1(\kappa), \tag{5}$$

$$\|v - \Pi_N^\kappa v\|_{\star,\kappa} \le \|v\|_{\star,\kappa}, \qquad \forall v \in H^1(\kappa).$$

2.3. Local scaling constants

In this section, we are going to define some local constants that will be used in the upcoming a posteriori error analysis. We start with defining the local trace inverse inequality constant d_κ for each $\kappa \in \mathcal{K}$ defined by

$$\mathsf{d}_\kappa \equiv \sup_{v_N \in \mathbb{V}_N(\kappa)} \frac{\|\nabla v_N \cdot n_\kappa\|_{\partial\kappa}}{\|v_N\|_{\star,\kappa}} > 0.$$

Further, let

$$\mathsf{a}_\kappa \equiv \sup_{\substack{v \in H^1(\kappa), \\ v \perp \mathbb{V}_N(\kappa)}} \frac{\|v\|_\kappa}{\|v\|_{\star,\kappa}} \qquad \text{and} \qquad \mathsf{b}_\kappa \equiv \sup_{\substack{v \in H^1(\kappa), \\ v \perp \mathbb{V}_N(\kappa)}} \frac{\|v\|_{\partial\kappa}}{\|v\|_{\star,\kappa}},$$

where $\perp$ is in the sense of the scalar product $(\cdot, \cdot)_{\star,\kappa}$ defined by (2).

Remark 2.2 (The computation of the constants a_κ, b_κ and d_κ). *We provide more details in Section 5 on how these local constants can be approximated by solving local eigenvalue problems.*

Lemma 2.3. *Let $\kappa \in \mathcal{K}$, $v \in H^1(\kappa)$. Then, there hods that*

$$\|v - \Pi_N^\kappa v\|_\kappa \le \mathsf{a}_\kappa \|\nabla v\|_\kappa,$$

$$\|v - \Pi_N^\kappa v\|_{\partial\kappa} \le \mathsf{b}_\kappa \|\nabla v\|_\kappa.$$

Proof. The proof consists of simply combining the definition of a_κ resp. b_κ and the stability of the projection Π_N^κ described in (5)

$$\|v - \Pi_N^\kappa v\|_\kappa \le \mathsf{a}_\kappa \|v - \Pi_N^\kappa v\|_{\star,\kappa} = \mathsf{a}_\kappa \|\nabla(v - \Pi_N^\kappa v)\|_\kappa \le \mathsf{a}_\kappa \|\nabla v\|_\kappa,$$

since $\Pi_0^\kappa(v - \Pi_N^\kappa v) = 0$. The proof for the second inequality is almost identical. $\square$

3. Poisson's equation

As has been motivated in the introduction we start with a simple model problem that however reflects the difficulties associated to the discontinuous Galerkin method using non-polynomial functions. The problem then reads: find $u \in H^1_\#(\Omega) \cap H^2(\mathcal{K})$ such that

$$-\Delta u = f, \quad \text{in } \Omega, \tag{6}$$

for some $f \in L^2(\Omega)$.

Given a piecewise constant and positive penalty function γ such that $\gamma|_\kappa = \gamma_\kappa \in \mathbb{R}^+$ for all $\kappa \in \mathcal{K}$, the discontinuous bilinear form is defined by

$$a(w,v) = \sum_{\kappa \in \mathcal{K}} \left[(\nabla w, \nabla v)_\kappa - \tfrac{1}{2}(\nabla w, [\![v]\!])_{\partial\kappa} - \tfrac{\theta}{2}([\![w]\!], \nabla v)_{\partial\kappa} + \tfrac{\gamma_\kappa}{2}([\![w]\!], [\![v]\!])_{\partial\kappa} \right],$$

for any $w, v \in H^2(\mathcal{K})$ and for $\theta \in \mathbb{R}$. Note that this is equivalent to the somewhat more standard notation

$$a(w,v) = (\nabla w, \nabla v)_\mathcal{K} - (\{\!\{\nabla w\}\!\}, [\![v]\!])_\mathcal{F} - \theta([\![w]\!], \{\!\{\nabla v\}\!\})_\mathcal{F} + (\gamma_\mathbf{F}[\![w]\!], [\![v]\!])_\mathcal{F},$$

with $\gamma_\mathbf{F} = \{\!\{\gamma\}\!\}$ and where $(\cdot, \cdot)_\mathcal{F}$ denotes the face-wise L^2-inner product over all faces of the mesh. Note that the choice of $\theta = 1, -1$ corresponds to the symmetric and non-symmetric interior penalty discontinuous Galerkin (SIPG [3, 27] or NIPG [6]) method, respectively. The former case results in a symmetric bilinear form.

Then, the **discontinuous Galerkin** approximation is defined by: Find $u_N \in \mathbb{V}_N$ such that

$$a(u_N, v_N) = (f, v_N)_\Omega, \qquad \forall v_N \in \mathbb{V}_N. \tag{7}$$

In this context we define the following **broken energy norm** by

$$\|v\|^2 = \sum_{\kappa \in \mathcal{K}} \left[\|\nabla v\|^2_\kappa + \tfrac{\gamma_\kappa}{2} \|[\![v]\!]\|^2_{\partial\kappa} \right], \quad \forall v \in H^1(\mathcal{K}). \tag{8}$$

Observe that

$$\|v\|^2 = \sum_{\kappa \in \mathcal{K}} \|v\|^2_\kappa \qquad \text{with} \qquad \|v\|^2_\kappa = \|\nabla v\|^2_\kappa + \tfrac{\gamma_\kappa}{2} \|[\![v]\!]\|^2_{\partial\kappa},$$

and that this is indeed a norm as $\gamma > 0$. As usual, the penalty parameter γ needs to be chosen carefully to ensure coercivity. Even when polynomial basis functions are used, the choice of an optimal γ is not completely trivial and related discussions can be found in [1,9]. The scaling in the element sizes and the polynomial order is however known [15, 25]. The involved constants are resulting from applying trace and inverse inequalities, but no inverse inequality is known if general non-polynomial basis functions are employed. To have a precise idea of the values of the combined trace and inverse inequalities for the generic non-polynomial basis functions spanning $\mathbb{V}_N$, we propose here to use the local constants that were defined in Section 2. In consequence, we can give a precise value of the piecewise constant function γ that is needed to ensure coercivity of the bilinear form $a(\cdot, \cdot)$. This is stated in the following lemma.

Lemma 3.1. *Under the assumption that $\|\cdot\|$ is a norm (which is assumed here since $\gamma > 0$) and if additionally $\gamma_\kappa \geq \frac{1}{2}(1+\theta)^2 (\mathtt{d}_\kappa)^2$ for each $\kappa \in \mathcal{K}$, then the bilinear form is coercive on $\mathbb{V}_N$, i.e., there holds*

$$\tfrac{1}{2}\|v_N\|^2 \leq a(v_N, v_N), \qquad \forall v_N \in \mathbb{V}_N.$$

Proof. Since for any $v_N \in \mathbb{V}_N$ we have $\nabla v_N = \nabla(v_N - \Pi_0^\kappa v_N)$ and $\|v_N - \Pi_0^\kappa v_N\|_{\star,\kappa} = \|\nabla v_N\|_\kappa$ we can develop

$$a(v_N, v_N) = \frac{1}{2} \sum_{\kappa \in \mathcal{K}} \left[2 \|\nabla v_N\|_\kappa^2 - (1+\theta)(\nabla(v_N - \Pi_0^\kappa v_N), [\![v_N]\!])_{\partial\kappa} + \gamma_\kappa \|[\![v_N]\!]\|_{\partial\kappa}^2 \right]$$

$$\geq \frac{1}{2} \sum_{\kappa \in \mathcal{K}} \left[2 \|\nabla v_N\|_\kappa^2 - (1+\theta)\, \mathsf{d}_\kappa \|v_N - \Pi_0^\kappa v_N\|_{\star,\kappa} \|[\![v_N]\!]\|_{\partial\kappa} + \gamma_\kappa \|[\![v_N]\!]\|_{\partial\kappa}^2 \right]$$

$$\geq \frac{1}{2} \sum_{\kappa \in \mathcal{K}} \left[\|\nabla v_N\|_\kappa^2 + \left(\gamma_\kappa - \frac{1}{4}(1+\theta)^2 (\mathsf{d}_\kappa)^2 \right) \|[\![v_N]\!]\|_{\partial\kappa}^2 \right]$$

and obtain

$$\frac{1}{2} |\!|\!| v_N |\!|\!|^2 \leq a(v_N, v_N)$$

for the particular choice $\gamma_\kappa \geq \frac{1}{2}(1+\theta)^2 (\mathsf{d}_\kappa)^2$. Note however that for the particular choice of $\theta = -1$, γ_κ stills needs to be positive in order that $|\!|\!| \cdot |\!|\!|$ is indeed a norm. $\qquad\square$

3.1. **Error representation**

Define the scaled error function $\varphi = \frac{u - u_N}{|\!|\!| u - u_N |\!|\!|}$ and develop

$$|\!|\!| u - u_N |\!|\!| = \sum_{\kappa \in \mathcal{K}} \left[(\nabla(u - u_N), \nabla\varphi)_\kappa + \frac{\gamma_\kappa}{2}([\![u - u_N]\!], [\![\varphi]\!])_{\partial\kappa} \right]$$

$$= a(u - u_N, \varphi) + \frac{1+\theta}{2} \sum_{\kappa \in \mathcal{K}} (\nabla\varphi, [\![u - u_N]\!])_{\partial\kappa}.$$

We prefer to work with the scaled error function φ for sake of a simple presentation of the upcoming error analysis. Observe that due to the regularity of $u \in H_\#^1(\Omega)$, which implies $[\![u]\!] = 0$, and since u is indeed the solution of (6), there holds

$$a(u, \varphi) = \sum_{\kappa \in \mathcal{K}} \left[(\nabla u, \nabla\varphi)_\kappa - \frac{1}{2}(\nabla u, [\![\varphi]\!])_{\partial\kappa} \right] = -(\Delta u, \varphi)_\Omega = (f, \varphi)_\Omega.$$

On the other hand, since $u_N \in \mathbb{V}_N$ is the DG-solution solution of (7), we obtain

$$-a(u_N, \varphi) = -a(u_N, \varphi - \varphi_N) - (f, \varphi_N)_\Omega,$$

for any $\varphi_N \in \mathbb{V}_N$. Thus, using the integration by parts and Lemma 2.1, we can develop

$$-a(u_N, \varphi) = \sum_{\kappa \in \mathcal{K}} \left[-(\nabla u_N, \nabla(\varphi - \varphi_N))_\kappa + \frac{1}{2}(\nabla u_N, [\![\varphi - \varphi_N]\!])_{\partial\kappa} + \frac{\theta}{2}([\![u_N]\!], \nabla(\varphi - \varphi_N))_{\partial\kappa} \right.$$

$$\left. - \frac{\gamma_\kappa}{2}([\![u_N]\!], [\![\varphi - \varphi_N]\!])_{\partial\kappa} \right] - (f, \varphi_N)_\Omega$$

$$= \sum_{\kappa \in \mathcal{K}} \left[(\Delta u_N, \varphi - \varphi_N)_\kappa - \frac{1}{2}([\![\nabla u_N]\!], \varphi - \varphi_N)_{\partial\kappa} + \frac{\theta}{2}([\![u_N]\!], \nabla(\varphi - \varphi_N))_{\partial\kappa} \right.$$

$$\left. - \frac{\gamma_\kappa}{2}([\![u_N]\!], [\![\varphi - \varphi_N]\!])_{\partial\kappa} \right] - (f, \varphi_N)_\Omega,$$

and obtain the **error representation equation**

$$|\!|\!| u - u_N |\!|\!| = \sum_{\kappa \in \mathcal{K}} \left[(f + \Delta u_N, \varphi - \varphi_N)_\kappa - \frac{1}{2}([\![\nabla u_N]\!], \varphi - \varphi_N)_{\partial\kappa} \right. \tag{9}$$

$$\left. - \gamma_\kappa([\![u_N]\!], (\varphi - \varphi_N)n_\kappa)_{\partial\kappa} - \frac{1}{2}([\![u_N]\!], \nabla\varphi + \theta\nabla\varphi_N)_{\partial\kappa} \right].$$

3.2. A posteriori error estimation

After recalling that we assumed that $u \in H^2(\kappa)$, we start by introducing the constant $\mathsf{d}_\kappa^u(u_N)$ defined by

$$\mathsf{d}_\kappa^u(u_N) = \frac{\|\nabla(u - u_N) \cdot n_\kappa\|_{\partial\kappa}}{\|\nabla(u - u_N)\|_\kappa},$$

and define the constant c_κ by

$$\mathsf{c}_\kappa = \mathsf{d}_\kappa^u(u_N) + \mathsf{d}_\kappa|\theta|.$$

We note that in practice, the constant $\mathsf{d}_\kappa^u(u_N)$ can not be evaluated since u is unknown. See in the upcoming numerical examples how we deal with this term.

Remark 3.2. *Observe that $\mathsf{d}_\kappa^u(u_N)$ is bounded by the constant*

$$\sup_{v_N \in \mathbb{V}_N(\kappa)} \frac{\|\nabla(u - v_N) \cdot n_\kappa\|_{\partial\kappa}}{\|\nabla(u - v_N)\|_\kappa} < \infty,$$

which, in turn, is independent of the approximation u_N (but still depends on the exact solution u and the approximation space $\mathbb{V}_N$).

Define the following estimators

$$\eta_{\mathrm{R},\kappa} \equiv \mathsf{a}_\kappa \|f + \Delta u_N\|_\kappa, \tag{10}$$

$$\eta_{\mathrm{F},\kappa} \equiv \tfrac{\mathsf{b}_\kappa}{2}\|[\![\nabla u_N]\!]\|_{\partial\kappa}, \tag{11}$$

$$\eta_{\mathrm{J},\kappa} \equiv (\mathsf{b}_\kappa\,\gamma_\kappa + \tfrac{\mathsf{c}_\kappa}{2})\|[\![u_N]\!]\|_{\partial\kappa}, \tag{12}$$

in order to state the first Theorem.

Theorem 3.3. *Let $u \in H^1_\#(\Omega) \cap H^2(\mathcal{K})$ be the solution of (6) and $u_N \in \mathbb{V}_N$ the DG-approximation defined by (7). Then, we have the following a posteriori upper bound*

$$\|u - u_N\| \leq \left(\sum_{\kappa \in \mathcal{K}} \left[\eta_{\mathrm{R},\kappa} + \eta_{\mathrm{F},\kappa} + \eta_{\mathrm{J},\kappa}\right]^2\right)^{\frac{1}{2}}.$$

Proof. Using the triangle inequality, observe that

$$\|(\nabla\varphi + \theta\nabla\varphi_N) \cdot n_\kappa\|_{\partial\kappa} \leq \|\nabla\varphi \cdot n_\kappa\|_{\partial\kappa} + |\theta|\|\nabla\varphi_N \cdot n_\kappa\|_{\partial\kappa} \leq \mathsf{d}_\kappa^u(u_N)\|\nabla\varphi\|_\kappa + \mathsf{d}_\kappa|\theta|\|\nabla\varphi_N\|_\kappa.$$

So far, the results where valid for any arbitrary discrete function $\varphi_N \in \mathbb{V}_N$. In this proof we consider the particular choice $\varphi_N|_\kappa = \Pi_N^\kappa \varphi$ so that we can easily state

$$\|\nabla\varphi_N\|_\kappa \leq \|\nabla\varphi\|_\kappa$$

by splitting $\varphi = \Pi_N^\kappa \varphi + (\varphi - \Pi_N^\kappa \varphi)$ and using the orthogonality relation (4). Then, there holds

$$\|(\nabla\varphi + \theta\nabla\varphi_N) \cdot n_\kappa\|_{\partial\kappa} \leq \underbrace{(\mathsf{d}_\kappa^u(u_N) + \mathsf{d}_\kappa|\theta|)}_{=\mathsf{c}_\kappa} \|\nabla\varphi\|_\kappa \leq \mathsf{c}_\kappa\|\nabla\varphi\|_\kappa, \tag{13}$$

by applying a simple triangle inequality.

If we apply the Cauchy-Schwarz inequality to the error representation formula (9), in combination with Lemma 2.3, Eq. (13) and another Cauchy-Schwarz inequality, we have (recall that $\varphi = \frac{u-u_N}{\|u-u_N\|}$)

$$
\begin{aligned}
\|u - u_N\| &\leq \sum_{\kappa \in \mathcal{K}} \Big[\|f + \Delta u_N\|_\kappa \|\varphi - \varphi_N\|_\kappa + \tfrac{1}{2} \|[\![\nabla u_N]\!]\|_{\partial\kappa} \|\varphi - \varphi_N\|_{\partial\kappa} + \gamma_\kappa \|[\![u_N]\!]\|_{\partial\kappa} \|\varphi - \varphi_N\|_{\partial\kappa} \\
&\qquad + \tfrac{1}{2} \|[\![u_N]\!]\|_{\partial\kappa} \|(\nabla\varphi + \theta\nabla\varphi_N)\cdot n_\kappa\|_{\partial\kappa} \Big] \\
&\leq \sum_{\kappa \in \mathcal{K}} \Big[\mathsf{a}_\kappa \|f + \Delta u_N\|_\kappa + \tfrac{\mathsf{b}_\kappa}{2} \|[\![\nabla u_N]\!]\|_{\partial\kappa} + (\gamma_\kappa \mathsf{b}_\kappa + \tfrac{\mathsf{c}_\kappa}{2}) \|[\![u_N]\!]\|_{\partial\kappa} \Big] \|\nabla\varphi\|_\kappa \\
&\leq \left(\sum_{\kappa \in \mathcal{K}} \Big[\mathsf{a}_\kappa \|f + \Delta u_N\|_\kappa + \tfrac{\mathsf{b}_\kappa}{2} \|[\![\nabla u_N]\!]\|_{\partial\kappa} + (\gamma_\kappa \mathsf{b}_\kappa + \tfrac{\mathsf{c}_\kappa}{2}) \|[\![u_N]\!]\|_{\partial\kappa} \Big]^2 \right)^{\frac{1}{2}} \\
&= \left(\sum_{\kappa \in \mathcal{K}} \Big[\eta_{\mathrm{R},\kappa} + \eta_{\mathrm{F},\kappa} + \eta_{\mathrm{J},\kappa} \Big]^2 \right)^{\frac{1}{2}}.
\end{aligned}
$$

$\square$

4. Second order indefinite problems

In this section we consider the more general indefinite equation: find $u \in H^1_\#(\Omega)$ such that

$$- \Delta u + V u = f, \qquad \text{in } \Omega, \tag{14}$$

for some $f \in L^2(\Omega)$ and where we only assume that $V \in L^\infty(\Omega)$ is bounded and that the operator $-\Delta + V$ has no zero eigenvalue. For the particular choice of $V = -k^2 \in \mathbb{R}$ this framework includes the Helmholtz equation. The DG-bilinear form is provided by

$$a(w,v) = \sum_{\kappa \in \mathcal{K}} \Big[(\nabla w, \nabla v)_\kappa + (Vw, v)_\kappa - \tfrac{1}{2}(\nabla w, [\![v]\!])_{\partial\kappa} - \tfrac{\theta}{2}([\![w]\!], \nabla v)_{\partial\kappa} + \tfrac{\gamma_\kappa}{2}([\![w]\!], [\![v]\!])_{\partial\kappa} \Big],$$

such that the DG-approximation is defined by: Find $u_N \in \mathbb{V}_N$ such that

$$a(u_N, v_N) = (f, v_N)_\Omega, \qquad \forall v_N \in \mathbb{V}_N, \tag{15}$$

and we keep the definition of the broken energy norm of (8). Of course the choice $\gamma_\kappa = 2(1 + |\theta|)^2 (\mathsf{d}_\kappa)^2$ does not imply coercivity of the bilinear form in this setting any more. We assume that γ_κ has been chosen by the user to insure that DG-problem has a unique solution and focus on how to quantify the error a posteriori. Observe that whenever the DG-problem is not uniquely solvable, the solver of the numerical system typically reveals the lack of well-posedness. The following analysis requires that the DG-solution satisfies (15).

4.1. Computable upper bounds

We first introduce a modified norm. For this consider V_+ and V_- defined by $V_+ = \max(V, 0) \geq 0$ and $V_- = \max(-V, 0) \geq 0$ so that $V = V_+ - V_-$ and $|V| = V_+ + V_-$. Then, define

$$\|v\|^2 = \sum_{\kappa \in \mathcal{K}} \|v\|_\kappa^2 \qquad \text{with} \qquad \|v\|_\kappa^2 = \|\nabla v\|_\kappa^2 + \|V_+^{\frac{1}{2}} v\|_\kappa^2 + \tfrac{\gamma_\kappa}{2} \|[\![v]\!]\|_{\partial\kappa}^2, \qquad \forall v \in H^1(\mathcal{K}).$$

Applying similar arguments as in Section 3 the following error representation can be developed

$$\|u - u_N\| = \sum_{\kappa \in \mathcal{K}} \Big[(f + \Delta u_N - V u_N, \varphi - \varphi_N)_\kappa + (V_-(u - u_N), \varphi)_\kappa \Big] \tag{16}$$

$$- \tfrac{1}{2} \sum_{\kappa \in \mathcal{K}} \Big[([\![\nabla u_N]\!], \varphi - \varphi_N)_{\partial\kappa} + \gamma_\kappa([\![u_N]\!], [\![\varphi - \varphi_N]\!])_{\partial\kappa} + ([\![u_N]\!], \nabla\varphi + \theta\nabla\varphi_N)_{\partial\kappa} \Big].$$

Redefining the residual as

$$\eta_{\mathsf{R},\kappa} \equiv \mathsf{a}_\kappa \|f + \Delta u_N - V u_N\|_\kappa, \tag{17}$$

the following bound can be developed.

Theorem 4.1. *Let $u \in H^1_\#(\Omega) \cap H^2(\mathcal{K})$ be the solution of (14) and $u_N \in \mathbb{V}_N$ the DG-approximation defined by (15). Then, we have the following a posteriori upper bound*

$$\|u - u_N\| \le \left(\sum_{\kappa \in \mathcal{K}} \left[\eta_{\mathsf{R},\kappa} + \eta_{\mathsf{F},\kappa} + \eta_{\mathsf{J},\kappa} \right]^2 \right)^{\frac{1}{2}} + \frac{\|V_-^{\frac{1}{2}}(u - u_N)\|_\mathcal{K}^2}{\|u - u_N\|},$$

where $\eta_{\mathsf{R},\kappa}$ is defined by (17) and $\eta_{\mathsf{F},\kappa}, \eta_{\mathsf{J},\kappa}$ are defined by (11)–(12).

Proof. This estimate can be obtained by applying the Cauchy-Schwarz inequality to the error representation equation (16) similar as in the proof of Theorem 3.3. Only the additional term

$$(V_-(u - u_N), \varphi)_\mathcal{K} = \frac{\|V_-^{\frac{1}{2}}(u - u_N)\|_\mathcal{K}^2}{\|u - u_N\|}$$

is not estimated. $\square$

Remark 4.2. *For $V_- \in L^\infty$, the term $\frac{\|V_-^{\frac{1}{2}}(u-u_N)\|_\mathcal{K}^2}{\|u-u_N\|}$ is small compared to the upper bound estimator in the limit of complete basis set. On the other hand, when only a small number of basis functions are used, this term can become large, and the upper bound error estimator can underestimate the true error in energy norm.*

4.2. Computable lower bounds

The goal of this section is to derive computable lower bounds of the approximation error. We note that the following theory applies also to the Poisson problem with $V = 0$.

Observe that

$$\eta_{\mathsf{J},\kappa} = \left(\mathsf{b}_\kappa \, \gamma_\kappa + \tfrac{\mathsf{c}_\kappa}{2} \right) \|[\![u_N]\!]\|_{\partial\kappa} \le \sqrt{\tfrac{2}{\gamma_\kappa}} \left(\mathsf{b}_\kappa \, \gamma_\kappa + \tfrac{\mathsf{c}_\kappa}{2} \right) \|u - u_N\|_\kappa.$$

Second, for any face F of $\partial\kappa$, denote by κ' the adjacent element such that $F = \partial\kappa \cap \partial\kappa'$ such that there holds

$$\eta_{\mathsf{F},\kappa}^2 = \tfrac{\mathsf{b}_\kappa^2}{4} \|[\![\nabla u_N]\!]\|_{\partial\kappa}^2 = \tfrac{\mathsf{b}_\kappa^2}{4} \|[\![\nabla(u - u_N)]\!]\|_{\partial\kappa}^2 \le \tfrac{\mathsf{b}_\kappa^2}{2} \sum_{F \in \partial\kappa} \left(\|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_F^2 + \|\nabla(u - u_N)|_{\kappa'} \cdot n_{\kappa'}\|_F^2 \right). \tag{18}$$

Recall that $\omega(\kappa)$ is the patch consisting of κ and its adjacent elements sharing one face, then

$$\eta_{\mathsf{F},\kappa}^2 \le \tfrac{\mathsf{b}_\kappa^2}{2} \sum_{\kappa' \in \omega(\kappa)} \left(\mathsf{d}_{\kappa'}^u(u_N) \|\nabla(u - u_N)\|_{\kappa'} \right)^2 \le \tfrac{\mathsf{b}_\kappa^2}{2} \left(\max_{\kappa' \in \omega(\kappa)} \mathsf{d}_{\kappa'}^u(u_N) \right)^2 \sum_{\kappa' \in \omega(\kappa)} \|\nabla(u - u_N)\|_{\kappa'}^2$$

Further, let g_κ be a smooth non-negative bubble function with $\sup_{x \in \kappa} g_\kappa(x) = 1$ and local support, i.e. $\mathrm{supp}(g_\kappa) \subset \kappa$, which in turn implies that $g_\kappa|_{\partial\kappa} = 0$. Finally, let us denote the residual by $R = f + \Delta u_N - V u_N$ and define

$$\sigma_\kappa = \mathsf{a}_\kappa \frac{\|R\|_\kappa}{\|g_\kappa^{\frac{1}{2}} R\|_\kappa^2}.$$

Denote by $\varphi_\kappa \in H_0^1(\kappa)$ the solution to

$$-\Delta\varphi_\kappa = V g_\kappa R, \qquad \text{on } \kappa,$$

so that

$$\eta_{\mathrm{R},\kappa} = \mathsf{a}_\kappa \|R\|_\kappa = \sigma_\kappa \|g_\kappa^{\frac{1}{2}} R\|_\kappa^2 = \sigma_\kappa \int_\kappa g_\kappa \Big[-\Delta(u - u_N) + V(u - u_N) \Big] R$$

$$= -\sigma_\kappa \int_\kappa \Big[\Delta(u - u_N)\, g_\kappa\, R - \Delta\varphi_\kappa(u - u_N) \Big]$$

$$= \sigma_\kappa \int_\kappa \Big[\nabla(u - u_N) \cdot \nabla(g_\kappa\, R) - \nabla(u - u_N) \cdot \nabla\varphi_\kappa \Big]$$

$$\leq \sigma_\kappa \|\nabla(u - u_N)\|_\kappa \|\nabla(g_\kappa\, R - \varphi_\kappa)\|_\kappa,$$

and in consequence

$$\frac{\eta_{\mathrm{R},\kappa}}{\|u - u_N\|_\kappa} \leq \sigma_\kappa \|\nabla(g_\kappa\, R - \varphi_\kappa)\|_\kappa.$$

The results above indicate that

$$\|u - u_N\|_\kappa \geq \max\left\{ \frac{\eta_{\mathrm{R},\kappa}}{c_{\mathrm{R},\kappa}}, \frac{\eta_{\mathrm{J},\kappa}}{c_{\mathrm{J},\kappa}} \right\}, \qquad \|u - u_N\|_{\omega(\kappa)} \geq \frac{\eta_{\mathrm{F},\kappa}}{c_{\mathrm{F},\kappa}}, \tag{19}$$

where, denoting by $|\omega(\kappa)|$ the cardinality of the set $\omega(\kappa)$, we use the definitions

$$\|v\|_{\omega(\kappa)}^2 = \frac{1}{|\omega(\kappa)|} \sum_{\kappa' \in \omega(\kappa)} \|\nabla v\|_{\kappa'}^2 + \tfrac{\gamma_\kappa}{2} \|[\![v]\!]\|_{\partial\kappa}^2,$$

and

$$c_{\mathrm{R},\kappa} = \mathsf{a}_\kappa \frac{\|R\|_\kappa \|\nabla(b_\kappa\, R - \varphi_\kappa)\|_\kappa}{\|b_\kappa^{1/2} R\|_\kappa^2},$$

$$c_{\mathrm{F},\kappa} = \mathsf{b}_\kappa \sqrt{\tfrac{|\omega(\kappa)|}{2}} \max_{\kappa' \in \omega(\kappa)} \mathsf{d}_{\kappa'}^u(u_N),$$

$$c_{\mathrm{J},\kappa} = \sqrt{\tfrac{2}{\gamma_\kappa}} \left(\mathsf{b}_\kappa\, \gamma_\kappa + \tfrac{c_\kappa}{2} \right).$$

We summarize the results in the following proposition.

Proposition 4.3 (Local lower bound). *Let $u \in H^1_\#(\Omega) \cap H^2(\mathcal{K})$ be the solution of (14) and $u_N \in \mathbb{V}_N$ the DG-approximation defined by (15). Then, the quantity*

$$\xi_\kappa = \max\left\{ \frac{\eta_{\mathrm{R},\kappa}}{c_{\mathrm{R},\kappa}}, \frac{\eta_{\mathrm{F},\kappa}}{c_{\mathrm{F},\kappa}}, \frac{\eta_{\mathrm{J},\kappa}}{c_{\mathrm{J},\kappa}} \right\},$$

is a local lower bound of the local error

$$\max\left\{ \|u - u_N\|_\kappa, \|u - u_N\|_{\omega(\kappa)} \right\}.$$

Remark 4.4. *Since in practice, the nominator as well as the denominator of any of those fractions might become very small, these ratios are not numerically stable. It turns out that*

$$\xi_\kappa = \frac{\eta_{\mathrm{R},\kappa} + \eta_{\mathrm{F},\kappa} + \eta_{\mathrm{J},\kappa}}{c_{\mathrm{R},\kappa} + c_{\mathrm{F},\kappa} + c_{\mathrm{J},\kappa}}$$

is numerically more robust and still meaningful as it replaces the maximum by the average.

On a global level, the following result holds.

Proposition 4.5 (Global lower bound)**.** *Let $u \in H^1_\#(\Omega) \cap H^2(\mathcal{K})$ be the solution of* (14) *and $u_N \in \mathbb{V}_N$ the DG-approximation defined by* (15)*. Then, there holds that*

$$\xi = \frac{\left(\sum_{\kappa \in \mathcal{K}} \left[\eta_{\mathsf{R},\kappa} + \eta_{\mathsf{F},\kappa} + \eta_{\mathsf{J},\kappa}\right]^2\right)^{\frac{1}{2}}}{\sqrt{3}\max_{\kappa \in \mathcal{K}}\left(c_{\mathsf{R},\kappa}^2 + \mathsf{b}_{\omega(\kappa)}^2 \mathsf{d}_\kappa^u(u_N)^2 + c_{\mathsf{J},\kappa}^2\right)^{\frac{1}{2}}} \leq \|u - u_N\|,$$

where

$$\mathsf{b}_{\omega(\kappa)}^2 = \max_{F \in \partial\kappa}\{\!\{\mathsf{b}_\kappa^2\}\!\}|_F = \max_{F \in \partial\kappa}\left(\frac{\mathsf{b}_\kappa^2}{2} + \frac{\mathsf{b}_{\kappa'}^2}{2}\right)\Big|_F.$$

Proof. Observe that by (18) there holds

$$\sum_{\kappa \in \mathcal{K}} \eta_{\mathsf{F},\kappa}^2 \leq \sum_{\kappa \in \mathcal{K}} \frac{\mathsf{b}_\kappa^2}{2} \sum_{F \in \partial\kappa} \|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_F^2 + \sum_{\kappa \in \mathcal{K}} \frac{\mathsf{b}_\kappa^2}{2} \sum_{F \in \partial\kappa} \|\nabla(u - u_N)|_{\kappa'} \cdot n_{\kappa'}\|_F^2$$

$$= \sum_{\kappa \in \mathcal{K}} \frac{\mathsf{b}_\kappa^2}{2} \sum_{F \in \partial\kappa} \|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_F^2 + \sum_{\kappa \in \mathcal{K}} \sum_{F \in \partial\kappa} \frac{\mathsf{b}_{\kappa'}^2}{2} \|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_F^2$$

$$= \sum_{\kappa \in \mathcal{K}} \sum_{F \in \partial\kappa} \left(\frac{\mathsf{b}_\kappa^2}{2} + \frac{\mathsf{b}_{\kappa'}^2}{2}\right) \|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_F^2 = \sum_{\kappa \in \mathcal{K}} \sum_{F \in \partial\kappa} \{\!\{\mathsf{b}_\kappa^2\}\!\} \|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_F^2$$

$$\leq \sum_{\kappa \in \mathcal{K}} \mathsf{b}_{\omega(\kappa)}^2 \|\nabla(u - u_N)|_\kappa \cdot n_\kappa\|_{\partial\kappa}^2 \leq \sum_{\kappa \in \mathcal{K}} \mathsf{b}_{\omega(\kappa)}^2 \mathsf{d}_\kappa^u(u_N)^2 \|\nabla(u - u_N)\|_\kappa^2.$$

Then, using the other local estimates for $\eta_{\mathsf{R},\kappa}$ and $\eta_{\mathsf{J},\kappa}$ given by (19) yields

$$\sum_{\kappa \in \mathcal{K}} \left[\eta_{\mathsf{R},\kappa} + \eta_{\mathsf{F},\kappa} + \eta_{\mathsf{J},\kappa}\right]^2 \leq 3\sum_{\kappa \in \mathcal{K}} \left(\eta_{\mathsf{R},\kappa}^2 + \eta_{\mathsf{F},\kappa}^2 + \eta_{\mathsf{J},\kappa}^2\right) \leq 3\sum_{\kappa \in \mathcal{K}} \left(c_{\mathsf{R},\kappa}^2 + \mathsf{b}_{\omega(\kappa)}^2 \mathsf{d}_\kappa^u(u_N)^2 + c_{\mathsf{J},\kappa}^2\right) \|u - u_N\|_\kappa^2$$

$$\leq 3\max_{\kappa \in \mathcal{K}} \left(c_{\mathsf{R},\kappa}^2 + \mathsf{b}_{\omega(\kappa)}^2 \mathsf{d}_\kappa^u(u_N)^2 + c_{\mathsf{J},\kappa}^2\right) \|u - u_N\|^2.$$

□

5. Practical strategies for estimating the constants

In this section we discuss how to compute the constants $\mathsf{d}_\kappa, \mathsf{a}_\kappa, \mathsf{b}_\kappa$ as defined in Section 2 in the a posteriori error estimator for general non-polynomial basis functions in the discontinuous Galerkin framework. The basic strategy is to discretize the infinite dimensional representative space $H^1(\kappa)$ using a finite dimensional space such as high order polynomials, and to replace the various inner products defined in Section 2 by discrete bilinear forms using Gauss quadrature. With the help of these bilinear forms, $\mathsf{d}_\kappa, \mathsf{a}_\kappa, \mathsf{b}_\kappa$ can be estimated by solving an eigenvalue problem, locally on each element κ.

5.1. **Finite dimensional discretization**

For simplicity let $\kappa = [0, h]^d, d = 1, 2, 3$ and all quantities be real. We start the discussion with $d = 1$, i.e. $\kappa = [0, h]$. All numerical quadrature are to be performed using the Legendre-Gauss-Lobatto (LGL) quadrature with N_g points. The LGL grid points are denoted by $\{y_j\}_{j=1}^{N_g}$, and the corresponding LGL weights by $\{\omega_j\}_{j=1}^{N_g}$. The Lobatto quadrature implies that

$$y_1 = 0, \quad y_{N_g} = h,$$

which facilitates the description of the boundary integrals as in the estimate of d_κ and b_κ. The LGL grid points $\{y_j\}_{j=1}^{N_g}$ correspond to a unique set of Lagrange polynomials of degree $(N_g - 1)$, denoted by $\{p_j(x)\}_{j=1}^{N_g}$, and satisfy

$$p_j(y_i) = \delta_{ij}, \quad 1 \leq i, j \leq N_g,$$

where δ_{ij} is the Kronecker δ function. We can then approximate $v \in H^1(\kappa)$ using the linear combination of Lagrange polynomials as

$$v(x) \approx \sum_{j=1}^{N_g} v_j \, p_j(x).$$

The sequence of spaces $\mathbb{P}_{N_g}$ of polynomials of degree N_g being dense in $H^1(\kappa)$ implies that, for any $v \in H^1(\kappa)$ and any $\varepsilon > 0$, there exists N_g and $v^1_{N_g}, v^2_{N_g} \in \mathbb{P}_{N_g}$ such that

$$\frac{\|v - v^1_{N_g}\|_\kappa}{\|v\|_\kappa} \leq \varepsilon \qquad \text{and similarly} \qquad \frac{\|v - v^2_{N_g}\|_{\star,\kappa}}{\|v\|_{\star,\kappa}} \leq \varepsilon,$$

if choosing N_g large enough. That is, elements in $H^1(\kappa)$ can be approximated, in the sense of L^2 and H^1 with any desired accuracy by elements of $\mathbb{P}_{N_g}$. This motivates us to work in $\mathbb{P}_{N_g}$ instead of $H^1(\kappa)$ for N_g large enough. We assume that N_g is large enough so that the above approximation error in the local L^2 and H^1-norms are very small. Further, for functions $u, v \in \mathbb{P}_{N_g}$, the LGL quadrature for computing the inner product $(u, v)_\kappa$ converges rapidly with respect to the increase of N_g.

We denote by $v = (v_1, \ldots, v_{N_g})^T$ the column vector corresponding to the coefficients of $v \in \mathbb{P}_{N_g}$, and denote by $Y = (y_1, \ldots, y_{N_g})^T$, $w = (\omega_1, \ldots, \omega_{N_g})^T$ the column vector corresponding to the LGL grid points and weights, respectively. With a slight abuse of notation we can compute the inner product using linear algebra notation as

$$(u, v)_\kappa = \sum_{j=1}^{N_g} u_j \omega_j v_j \equiv u^T W v, \tag{20}$$

where $W = \text{diag}[w]$ is a diagonal matrix with the entries of vector w on the diagonal entries.

The Lagrange polynomials also induce a *differentiation matrix* D of size $N_g \times N_g$, defined as

$$D_{ij} = p'_j(y_i), \quad 1 \leq i, j \leq N_g. \tag{21}$$

Taking the derivative of a polynomial yields

$$v'(x) = \sum_{j=1}^{N_g} p'_j(x) v_j.$$

Let $v' = (v'(y_1), \ldots, v'(y_{N_g}))^T$ be the column vector of the derivative quantity $v'(x)$ on the LGL grid points, then

$$v' = Dv. \tag{22}$$

Eq. (22) shows that the differentiation matrix maps the values of a function to the values of its derivative on the LGL grid points. Using the differentiation matrix, inner products of the form $(u', v')_\kappa$ can be expressed in linear algebra notation as

$$(u', v')_\kappa = (Du)^T W (Dv) = u^T (D^T W D) v. \tag{23}$$

In order to compute the inner product $(u, v)_{\star,\kappa}$ we also need to compute $(\Pi_0^\kappa u, \Pi_0^\kappa v)_\kappa$. Note that

$$\Pi_0^\kappa v = \frac{1}{|\kappa|}(1, v)_\kappa = \frac{1}{|\kappa|} w^T v,$$

with $|\kappa| = h$. Then

$$(\Pi_0^\kappa u, \Pi_0^\kappa v)_\kappa = \frac{1}{|\kappa|^2} u^T w w^T v |\kappa| = u^T \left(w \frac{1}{|\kappa|} w^T \right) v.$$

Therefore the inner product $(u, v)_{\star,\kappa}$ can be computed as

$$(u, v)_{\star,\kappa} = u^T \left(D^T W D + w \frac{1}{|\kappa|} w^T \right) v. \tag{24}$$

We also need to compute inner products on the boundary $\partial\kappa$. In 1D, $v|_{\partial\kappa}(x)$ is completely described by two points $v(0)$ and $v(h)$, which is given by the discretization on the LGL grid points as v_1 and v_{N_g}. Define the weight vector at 0-dimension as $\widetilde{w} = (1, 0, \ldots, 0, 1)^T$, and $\widetilde{W} = \mathrm{diag}[\widetilde{w}]$, then the inner product on the boundary can be expressed as

$$(u, v)_{\partial\kappa} = u_1 v_1 + u_{N_g} v_{N_g} \equiv u^T \widetilde{W} v. \tag{25}$$

Similarly

$$(u', v')_{\partial\kappa} = u'_1 v'_1 + u'_{N_g} v'_{N_g} \equiv u^T D^T \widetilde{W} D v. \tag{26}$$

The inner products (20), (23), (24) and (26) are sufficient for estimating $\mathsf{d}_\kappa, \mathsf{a}_\kappa, \mathsf{b}_\kappa$ for $d = 1$.

Now we generalize all the definition above to $d > 1$. Though in practice we only consider $d = 2, 3$, the formalism developed here holds for any dimension. For any $x \in \kappa = [0, h]^d$, we denote by $x = (x^{(1)}, \ldots, x^{(d)})^T$, with $x^{(l)}$ being the component of x along the l-th dimension. Then the set of N_g^d LGL grid points in the dimension d is given by

$$Y^{[d]} = \{y_{j_1, \ldots, j_d} \equiv (y_{j_1}, \ldots, y_{j_d})^T | 1 \leq j_1, \ldots, j_d \leq N_g\}. \tag{27}$$

We define the tensor product of d matrices $A^{(1)}, \ldots, A^{(d)}$ of size $N_g \times N_g$ as

$$A_{i_1 j_1, \ldots, i_d j_d} = \prod_{l=1}^d A_{i_l j_l}^{(l)}, \quad 1 \leq i_1, j_1, \ldots, i_d, j_d \leq N_g, \tag{28}$$

which can be written in a compact form as

$$A \equiv \bigotimes_{l=1}^d A^{(l)}. \tag{29}$$

From the computational point of view, it is more convenient to rewrite the tensor product A as a matrix by stacking the $i_1, \ldots, i_d$ and $j_1, \ldots, j_d$ indices, respectively. In other words, we can view A as a large matrix of size $N_g^d \times N_g^d$, and each matrix element $A_{i_1 j_1, \ldots, i_d j_d}$ corresponds to a matrix element $A_{\mathcal{I}\mathcal{J}}$, with the index

$$\mathcal{I} = 1 + \sum_{l=1}^d (i_l - 1) N_g^{(l-1)}, \quad \mathcal{J} = 1 + \sum_{l=1}^d (j_l - 1) N_g^{(l-1)}.$$

Note that when $d = 2$, the stacked representation of the tensor product of $A^{(1)}$ and $A^{(2)}$ is the Kronecker product of $A^{(2)}$ and $A^{(1)}$.

We also define a special case for the tensor product of d vectors $v^{(1)}, \ldots, v^{(d)}$ of size N_g. By viewing each $v^{(l)}$ as a matrix of size $N_g \times 1$, we have

$$v_{j_1, \ldots, j_d} = \prod_{l=1}^d v_{j_l}^{(l)}, \quad 1 \leq j_1, \ldots, j_d \leq N_g. \tag{30}$$

Eq. (30) can be written in a compact form as

$$v \equiv \bigotimes_{l=1}^d v^{(l)}. \tag{31}$$

By stacking the indices $j_1, \ldots, j_d$ together, we can view v as a vector of size N_g^d, and each element $v_{j_1, \ldots, j_d}$ corresponds to an element $v_{\mathcal{J}}$ with $\mathcal{J} = 1 + \sum_{l=1}^d (j_l - 1) N_g^{(l-1)}$. Using the notation of tensor product, the set of LGL weights is described by a vector

$$w^{[d]} = \bigotimes_{l=1}^d w. \tag{32}$$

Similar to the 1D case, each LGL grid point $y_{j_1,\ldots,j_d}$ uniquely corresponds to a Lagrange polynomial

$$p_{j_1,\ldots,j_d}(x) = \prod_{l=1}^{d} p_{j_l}(x^{(l)}).$$

It can be readily seen that

$$p_{j_1,\ldots,j_d}(y_{i_1,\ldots,i_d}) = \prod_{l=1}^{d} \delta_{i_l j_l}.$$

As in the 1D case, a polynomial $u(x)$ defined on κ can be expressed using the Lagrange polynomials as

$$u(x) = \sum_{1 \leq j_1,\ldots,j_d \leq N_g} p_{j_1,\ldots,j_d}(x) u(y_{j_1,\ldots,j_d}) \equiv \sum_{1 \leq j_1,\ldots,j_d \leq N_g} p_{j_1,\ldots,j_d}(x) u_{j_1,\ldots,j_d}. \tag{33}$$

Denote by $W^{[d]} = \mathrm{diag}[w^{[d]}]$ as a matrix of size $N_g^d \times N_g^d$, the inner product $(u, v)_\kappa$ can be written as

$$(u, v)_\kappa = \sum_{1 \leq j_1,\ldots,j_d \leq N_g} u_{j_1,\ldots,j_d} v_{j_1,\ldots,j_d} w^{[d]}_{j_1,\ldots,j_d} = u^T W^{[d]} v. \tag{34}$$

The Lagrange polynomials $p_{j_1,\ldots,j_d}(x)$ can be used to define d differentiation matrices, defined as

$$D_l^{[d]} = \left(\bigotimes_{i=1}^{l-1} I \right) \bigotimes D \bigotimes \left(\bigotimes_{i=l+1}^{d} I \right). \tag{35}$$

Here I is an $N_g \times N_g$ identity matrix. $D_l^{[d]}$ can be understood as the discretized differential operator $\partial_l, 1 \leq l \leq d$. Similar to Eq. (22), we denote by $\partial_l v$ a column vector with its entries defined as below

$$(\partial_l v)_{j_1,\ldots,j_d} = (\partial_l v)(y_{i_1,\ldots,i_d}),$$

then $\partial_l v$ can be expressed in the linear algebra notation as

$$\partial_l v = D_l^{[d]} v. \tag{36}$$

Therefore the inner product $(\nabla u, \nabla v)_\kappa$ can be computed as

$$(\nabla u, \nabla v)_\kappa = u^T \left(\sum_{l=1}^{d} (D_l^{[d]})^T W^{[d]} D_l^{[d]} \right) v. \tag{37}$$

The inner product $(u, v)_{\star,\kappa}$ can be evaluated similar to Eq. (24) as

$$(u, v)_{\star,\kappa} = u^T \left(\sum_{l=1}^{d} (D_l^{[d]})^T W^{[d]} D_l^{[d]} + w^{[d]} \frac{1}{|\kappa|} (w^{[d]})^T \right) v, \tag{38}$$

with $|\kappa| = h^d$.

In order to evaluate the inner product on the boundary $\partial\kappa$, we define d weight vectors corresponding to the $(d-1)$ dimensional surface for each dimension l ($l = 1, \ldots, d$), denoted by $\widetilde{w}_l^{[d]}$ with the expression

$$\left(\widetilde{w}_l^{[d]} \right)_{j_1,\ldots,j_d} = \begin{cases} w^{[d-1]}_{j_1,\ldots,j_{l-1},j_{l+1},\ldots,j_d}, & j_l = 1 \quad \text{or} \quad j_l = N_g, \\ 0, & 1 < j_l < N_g. \end{cases} \tag{39}$$

Define $\widetilde{W}_l^{[d]} = \mathrm{diag}\left[\widetilde{w}_l^{[d]} \right]$, then the inner product on the boundary can be expressed as

$$(u, v)_{\partial\kappa} = u^T \left(\sum_{l=1}^{d} \widetilde{W}_l^{[d]} \right) v, \tag{40}$$

and

$$(\nabla u \cdot n_\kappa, \nabla v \cdot n_\kappa)_{\partial \kappa} = u^T \left(\sum_{l=1}^{d} (D_l^{[d]})^T \widetilde{W}_l^{[d]} D_l^{[d]} \right) v. \tag{41}$$

Now we are ready to use the finite dimensional representation of the inner products to evaluate the constants $d_\kappa, a_\kappa, b_\kappa$.

5.2. Estimation of d_κ

Recall that

$$(d_\kappa)^2 = \sup_{v_N \in \mathbb{V}_N(\kappa)} \frac{\|\nabla v_N \cdot n_\kappa\|_{\partial \kappa}^2}{\|v_N\|_{\star,\kappa}^2} \equiv \sup_{v_N \in \mathbb{V}_N(\kappa)} \frac{(\nabla v_N \cdot n_\kappa, \nabla v_N \cdot n_\kappa)_{\partial \kappa}}{(v_N, v_N)_{\star,\kappa}}.$$

Using Eq. (41) and Eq. (38), we have

$$(d_\kappa)^2 = \sup_{v_N \in \mathbb{V}_N(\kappa)} \frac{v_N^T M_\delta v_N}{v_N^T K v_N}. \tag{42}$$

Here

$$M_\delta = \sum_{l=1}^{d} (D_l^{[d]})^T \widetilde{W}_l^{[d]} D_l^{[d]}, \tag{43}$$

$$K = \sum_{l=1}^{d} (D_l^{[d]})^T W^{[d]} D_l^{[d]} + w^{[d]} \frac{1}{|\kappa|} (w^{[d]})^T. \tag{44}$$

Let $\{\varphi_1(x), \ldots, \varphi_N(x)\}$ be a set of basis functions of the finite dimensional space $\mathbb{V}_N(\kappa)$. We denote by $\varphi_i (i = 1, \ldots, N)$ the column vector corresponding to the values of $\varphi_i(x)$ evaluated at the LGL grid points, and denote by

$$\Phi = [\varphi_1, \ldots, \varphi_N], \tag{45}$$

the collection of all column vectors which is an $N_g^d \times N$ matrix. Then for any vector $v(x) \in \mathbb{V}_N(\kappa)$, the corresponding column vector v can be represented as

$$v = \Phi c,$$

where c is a coefficient vector of size N. Then Eq. (42) can be rewritten as

$$(d_\kappa)^2 = \sup_{c \in \mathbb{R}^N} \frac{c^T (\Phi^T M_\delta \Phi) c}{c^T (\Phi^T K_\delta \Phi) c}. \tag{46}$$

Eq. (46) can be solved as an eigenvalue problem,

$$\Phi^T M_\delta \Phi c = \lambda \Phi^T K_\delta \Phi c, \tag{47}$$

and $(d_\kappa)^2$ is equal to the largest eigenvalue λ. Since the size of the matrix $\Phi^T M_\delta \Phi$ is $N \times N$ and N is relatively small, Eq. (47) can be solved as a generalized eigenvalue problem using dense linear algebra.

5.3. Estimation of a_κ, b_κ

Recall that

$$a_\kappa^2 = \sup_{\substack{v \in H^1(\kappa), \\ v \perp \mathbb{V}_N(\kappa)}} \frac{\|v\|_\kappa^2}{\|v\|_{\star,\kappa}^2} = \sup_{\substack{v \in H^1(\kappa), \\ v \perp \mathbb{V}_N(\kappa)}} \frac{(v, v)_\kappa}{(v, v)_{\star,\kappa}},$$

then using Eq. (34), Eq. (38) and the density arguments shown above, it can be shown that

$$\sup_{\substack{v \in \mathbb{P}_{N_g}, \\ v \perp \mathbb{V}_N(\kappa)}} \frac{v^T M_a v}{v^T K v} \xrightarrow{N_g \to \infty} a_\kappa^2, \tag{48}$$

where $M_{\mathsf{a}} = W^{[d]}$, and K is given in Eq. (44). We can express the orthogonality condition $v \perp \mathbb{V}_N(\kappa)$ in terms of a projection operator $Q = \mathbb{I} - \Pi_N^\kappa$ so that for any $v \in H^1(\kappa)$, $Qv \perp \mathbb{V}_N(\kappa)$, where $\mathbb{I}$ is the identity operator. Denoting by Φ as in Eq. (45) the collection of spanning vectors of the space $\mathbb{V}_N(\kappa)$, then using the Lagrange polynomials corresponding to the LGL grid points as a basis, the projection operator Π_N^κ can be expressed as an $N_g^d \times N_g^d$ matrix

$$\Pi_N^\kappa = \Phi(\Phi^T K \Phi)^{-1} \Phi^T K \equiv \Phi \Psi^T. \tag{49}$$

where $\Psi = K\Phi(\Phi^T K \Phi)^{-1}$. Therefore the Π_N^κ is a low rank matrix with rank N. The projection operator Q and its adjoint operator Q^T expressed in the basis of Lagrange polynomials become

$$Q = I - \Phi\Psi^T, \quad Q^T = I - \Psi\Phi^T. \tag{50}$$

Using Eq. (50), the computation of a_κ can be simplified as

$$\mathsf{a}_\kappa^2 \approx \sup_{v \in \mathbb{R}^{N_g^d}} \frac{v^T Q^T M_{\mathsf{a}} Q v}{v^T Q^T K Q v}. \tag{51}$$

In other words, a_κ^2 corresponds to the largest eigenvalue of the generalized eigenvalue problem

$$Q^T M_{\mathsf{a}} Q v = \lambda Q^T K Q v. \tag{52}$$

From a computational point of view, there are two major differences between Eq. (47) and (52). First, the dimension of the matrices in Eq. (47) is $N \times N$, and the dimension of the matrices in Eq. (52) is $N_g^d \times N_g^d$. For 3D simulation, if $N_g = 30$ then $N_g^d = 27000$, and the corresponding eigenvalue problem is very costly to solve if $Q^T M_{\mathsf{a}} Q$ and $Q^T K Q$ are treated as dense matrices. Second, the matrix $\Phi^T K \Phi$ in Eq. (47) is a positive definite matrix since K is positive definite, and the problem (47) can be solved directly as a dense generalized eigenvalue problem. On the other hand, $Q^T K Q$ is a rank deficient matrix with the rank of its kernel being N. Therefore it can potentially cause a large numerical error if Eq. (52) is solved directly as a dense generalized eigenvalue problem.

In order to overcome the two difficulties mentioned above, we note that for any vector v, the computational cost for the matrix vector multiplication $Qv, Q^T v, M_{\mathsf{a}} v, K v$ is only proportional to N_g^d thanks to the low rank representation of the operators. Therefore Eq. (52) can be solved using iterative methods. Another advantage of using iterative methods is that since we only need the largest eigenvalue corresponding to Eq. (52), at the k-th step of the CG iteration we only need to keep three vectors: the current approximation of eigenvector $v^{(k)}$, the conjugate direction $p^{(k)}$ and the residual $r^{(k)}$. Even though the matrix $Q^T K Q$ is singular, the projection onto the 3 dimensional subspace $[v^{(k)}, p^{(k)}, r^{(k)}]$ is usually well conditioned. In practice we use the Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG) method [19] (with block size equal to 1) for evaluating the largest eigenvalue for Eq. (52). It should be noted that since there is no apparent preconditioner that can be applied efficiently to solve Eq. (52), the convergence of the largest eigenvalue may be slow. However, we should keep in mind that the estimation of $\mathsf{a}_\kappa, \mathsf{b}_\kappa$ is only used in the a posteriori error estimator, and only low accuracy is needed. In fact $\mathsf{a}_\kappa, \mathsf{b}_\kappa$ is already very accurate in the sense of the preconstant in the estimator even if the relative error is 10%. Therefore the slow convergence of the conjugate gradient method is compensated by the low accuracy required in the computation of the constants.

The constant b_κ can be estimated similarly to a_κ. Recall that

$$\mathsf{b}_\kappa^2 = \sup_{\substack{v \in H^1(\kappa), \\ v \perp \mathbb{V}_N(\kappa)}} \frac{\|v\|_{\partial\kappa}^2}{\|v\|_{\star,\kappa}^2} = \sup_{\substack{v \in H^1(\kappa), \\ v \perp \mathbb{V}_N(\kappa)}} \frac{(v,v)_{\partial\kappa}}{(v,v)_{\star,\kappa}},$$

and using the same projection operator Q, b_κ can be expressed as

$$\mathsf{b}_\kappa^2 \approx \sup_{v \in \mathbb{R}^{N_g^d}} \frac{v^T Q^T M_{\mathsf{b}} Q v}{v^T Q^T K Q v}, \tag{53}$$

with $M_{\mathsf{b}} = \sum_{l=1}^{d} \widetilde{W}_{l}^{[d]}$. Similar to Eq. (52), b_{κ}^{2} can be solved as the largest eigenvalue of

$$Q^{T} M_{\mathsf{b}} Q v = \lambda Q^{T} K Q v. \tag{54}$$

Eq. (54) can be solved using the same iterative strategy as for obtaining a_{κ}.

6. Numerical results

In this section we test the effectiveness of the a posteriori error estimators. The test program is written in MATLAB, and all results are obtained on a 2.7 GHz Intel processor with 16 GB memory. All numerical results are performed using the symmetric bilinear form ($\theta = 1$). The effectiveness of the upper bound and lower bound on the global domain will be justified by comparing $\|u - u_N\|$ and η, and by comparing $\|u - u_N\|$ and ξ, respectively. It should be noted that although our theory does not directly predict the effectiveness of the estimator on each local element κ, we can measure the local effectiveness of the upper and lower bound on each local element κ by defining

$$C_{\eta}(\kappa) = \frac{\eta_{\mathsf{R},\kappa} + \eta_{\mathsf{F},\kappa} + \eta_{\mathsf{J},\kappa}}{\|u - u_N\|_{\kappa}}, \quad C_{\xi}(\kappa) = \frac{\xi_{\kappa}}{\|u - u_N\|_{\kappa}}, \tag{55}$$

where the broken energy norm $\|u - u_N\|_{\kappa}$ is defined according to Eq. (8).

The numerical results are organized as follows. In section 6.1, we apply the general approach developed in section 5 to compute the constants $\mathsf{a}_{\kappa}, \mathsf{b}_{\kappa}, \mathsf{d}_{\kappa}$ for polynomial basis functions, and verify that the scaling properties of the numerically computed constants match the analytic results known in the literature [24]. In section 6.2, we illustrate the behavior of the upper bound and the lower bound error estimates for second order PDEs associated with positive definite operators. We then demonstrate the results for indefinite operators in section 6.3. In the a posteriori error estimates of both the upper bound and the lower bound, we make the assumption that the non-computable number $\mathsf{d}_{\kappa}^{u}(u_N)$ can be approximated by d_{κ} without significant loss of effectiveness. We justify such treatment in section 6.4 by directly calculating $\mathsf{d}_{\kappa}^{u}(u_N)$ using the numerically computed reference solution.

Our test problems include both one dimensional (1D) and two dimensional (2D) domains with periodic boundary conditions. Our non-polynomial basis functions are generated from the adaptive local basis (ALB) set [20] in the DG framework. The ALB set was originally proposed to systematically reduce the number of basis functions used to solve Kohn-Sham density functional theory calculations, and in this section we demonstrate its usage to solve second order linear PDEs. We denote by N the number of ALBs per element. For operators in the form of $A = -\Delta + V$ with periodic boundary condition, the basic idea of the ALB set is to use eigenfunctions computed local domains as basis functions corresponding to the lowest few eigenvalues. The eigenfunctions are associated with the same operator A, but with modified boundary conditions on the local domain. More specifically, in a d-dimensional space, for each element κ, we form an *extended element* $\widetilde{\kappa}$ consisting of κ and its $3^d - 1$ neighboring elements in the sense of periodic boundary condition. On $\widetilde{\kappa}$ we solve the eigenvalue problem

$$-\Delta \widetilde{\varphi}_i + V \widetilde{\varphi}_i = \lambda_i \widetilde{\varphi}_i. \tag{56}$$

with periodic boundary condition on $\partial \widetilde{\kappa}$. This eigenvalue problem can be solved using standard basis set such as finite difference, finite elements, or planewaves. Here we solve the local eigenvalue problem (56) using planewaves which naturally satisfy periodic boundary conditions. Since this eigenvalue problem is solved on a extended element $\widetilde{\kappa}$ the computational cost is not large. The collection of eigenfunctions (corresponding to lowest N eigenvalues) are restricted from $\widetilde{\kappa}$ to κ, i.e.

$$\varphi_i(x) = \begin{cases} [\widetilde{\varphi}_i]|_{\kappa}(x), & x \in \kappa; \\ 0, & \text{otherwise.} \end{cases}$$

After orthonormalizing $\{\varphi_i\}$ locally on each element κ and removing the linearly dependent functions, the resulting set of orthonormal functions are called the ALB functions.

Since periodic boundary condition is used on the global domain Ω, in all the calculations, the reference solution, which can be treated as a numerically exact solution, is solved using a planewave basis set with

a sufficiently large number of planewaves. The ALB basis set is also computed using a sufficiently large number of planewaves on the extended element $\widetilde{\kappa}$. Then a Fourier interpolation procedure is carried out from $\widetilde{\kappa}$ to the local element κ on a Legendre-Gauss-Lobatto (LGL) for accurate numerical integration.

6.1. **Estimating the constants for polynomial basis functions**

Although the main purpose of this paper is to design a posteriori error estimator for non-polynomial basis functions, the computational strategies discussed in section 5 can be applied to polynomial functions as well. Let $\kappa = [0, h]^d$ and $\mathbb{V}_N(p; \kappa) = \text{span}\{\prod_{l=1}^d x_l^{j_l}, j_l \in \mathbb{N}, \sum_{l=1}^d j_l \leq p\}$ be the space spanned by polynomials with degree less than or equal to p. Then the asymptotic scaling of $\mathsf{a}_\kappa, \mathsf{b}_\kappa, \mathsf{d}_\kappa$ with respect to h and p is known [14]

$$\mathsf{a}_\kappa^2 \sim \frac{h^2}{p^2}, \quad \mathsf{b}_\kappa^2 \sim \frac{h}{p}, \quad \mathsf{d}_\kappa^2 \sim \frac{p^2}{h}. \tag{57}$$

These results are asymptotically correct as $p \to \infty$, and we will show that the strategy discussed in section 5 leads to the same asymptotic result, but the result is more accurate in the pre-asymptotic regime due to the explicit computation of the constants.

From numerical point of view, the scaling with respect to h is naturally satisfied. To verify this, we can simply consider a reference element $\kappa|_{h=1} = [0, 1]^d$ and scale the weight matrix $W^{[d]}$ and the differentiation matrix $D_l^{[d]}$ accordingly. The technique is the same as that used in [24].

We now directly verify the scaling with respect to p in Fig. 1, using the algorithms presented in section 5. The LGL grid sizes for 1D, 2D and 3D calculation are chosen to be 100, 100×100, and $50 \times 50 \times 50$, respectively. The largest degree of polynomials is 64 for 1D and 2D, and is 16 for the 3D case. Note that in the 3D case, the dimension of $\mathbb{V}_N(p = 16; \kappa)$ is already 969. Fig. 1 (a) shows the behavior of a_κ^2, which asymptotically agrees with the $1/p^2$ scaling. It is interesting to see that the computed a_κ^2 can be approximated by $C\frac{h^2}{p^2}$ where the constant C is around 0.1. The recovery of the constant indicates that the numerically computed constant a_κ can offer a sharper estimator even for the standard hp-refinement. Similarly Fig. 1 (b) shows that b_κ^2 asymptotically scales as $1/p$ for 2D and 3D simulation. The 1D case is not shown in the picture, since the numerical value of b_κ^2 is already as small as 10^{-20} for $p = 2$. This can be interpreted from Proposition 7.1 in the appendix. Finally, direct computation in Fig. 1 (c) shows that d_κ^2 asymptotically scales as p^2 for all dimensions. Again, the computed constant d_κ^2 differs from the asymptotic scaling in the pre-asymptotic regime, indicating that the numerically computed constant should be sharper for low order polynomials ($p \leq 4$).
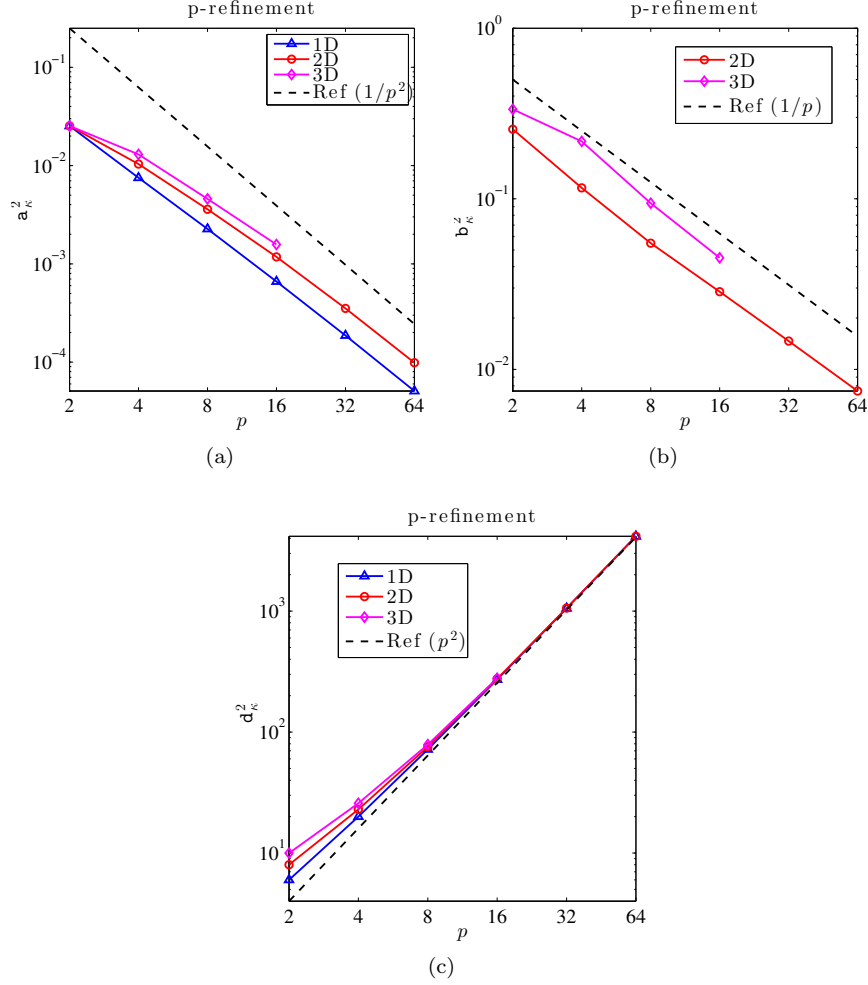
FIGURE 1. Numerically computed constants $\mathtt{a}_\kappa^2, \mathtt{b}_\kappa^2, \mathtt{d}_\kappa^2$ with respect to the polynomial degrees p in 1D, 2D and 3D.

6.2. Positive definite operators

We first demonstrate the effectiveness of the a posteriori error estimates for a positive definite operator on a 1D domain $\Omega = (0, 2\pi)$, using the ALB set as non-polynomial basis functions. Due to the periodic boundary condition, we choose $V(x) = 0.01$ so that the operator $A = -\Delta + V$ is non-singular and positive definite. The right hand side is chosen to be $f(x) = \sin(6x)$ which is periodic on Ω. In the ALB computation, the domain is partitioned into 7 elements, as indicated by black dashed lines. Fig. 2 shows solution u to Eq. (14) and the point-wise error $u - u_N$ using $N = 11$ ALBs per element.

Fig. 3 (a) shows the absolute error in the energy norm, the upper bound and lower bound estimates as the number of ALBs per element N increases from 3 to 15. The relative error can be deduced by comparing Fig. 3 (a) and Fig. 2 (a). We find that the computed η and ξ are indeed upper and lower bounds of the true error $\|u - u_N\|$ for all N across a wide range of accuracy (from 10^{-1} to 10^{-8}). It also appears that the lower bound estimator ξ follows the true error more closely than the upper bound estimator η. Fig. 3 (b) and (c) illustrate the local effectiveness $C_\eta(\kappa)$ and $C_\xi(\kappa)$ for each element κ. Though not guaranteed by our theory, we observe that η_κ and ξ_κ are upper and lower bounds for $\|u - u_N\|_\kappa$ for each element κ, respectively. The effectiveness as measured by $C_\eta(\kappa)$ and $C_\xi(\kappa)$ depends only weakly on the number of adaptive local basis functions, or the accuracy of the numerical solution.

Our next example is to solve a 2D problem with $\Omega = (0, 2\pi) \times (0, 2\pi)$. Again we choose $V(x, y) = 0.01$ so that $A = -\Delta + V$ is non-singular and positive definite. The right hand side is $f(x, y) = \cos(3x)\cos(y)$,
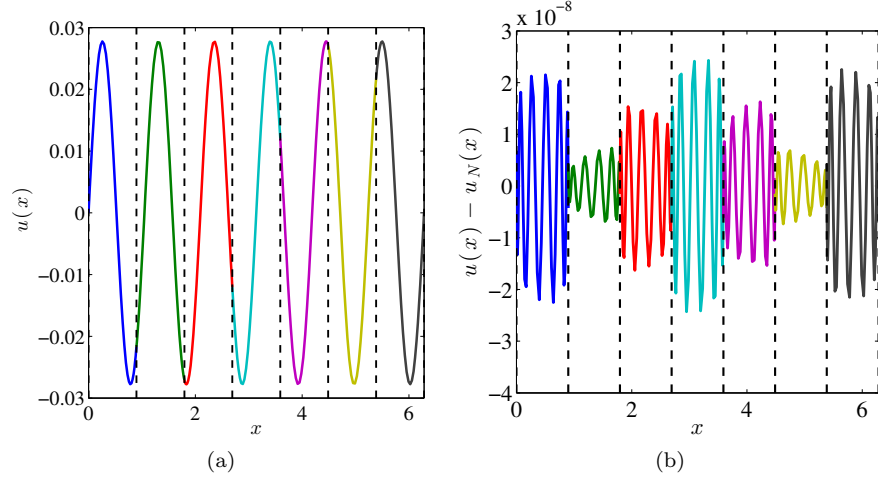
(a)   (b)

FIGURE 2. (a) The reference solution $u(x)$ corresponding to $V(x) = 0.01$ and the right
hand side $f(x) = \sin(6x)$. (b) Point-wise error between the reference solution $u(x)$ and
the numerical solution $u_N(x)$ calculated using the ALB set with 7 elements and $N = 11$
basis functions per element. The domain is partitioned into 7 elements indicated by
black dashed lines.

which satisfies the periodic boundary condition. Fig. 4 shows the reference solution u to Eq. (14) and
the point-wise error $u - u_N$ using $N = 31$ ALBs per element. In the ALB computation, the domain is
partitioned into 5×5 elements, indicated by black dashed lines.

    Fig. 5 (a) shows the error in the energy norm, the computed upper bound and the lower bound as
the number of ALBs per element N increases from 11 to 41. Both the computed upper and the lower
bound estimates are effective for all calculations. Fig. 5 (b)-(d) illustrates the local effectiveness of the
upper and lower bound estimates for the two extreme cases $N = 11$ and $N = 41$, and the estimator
η_κ and ξ_κ are effective for all elements, and the effectiveness depends weakly on the number of basis
functions per element.
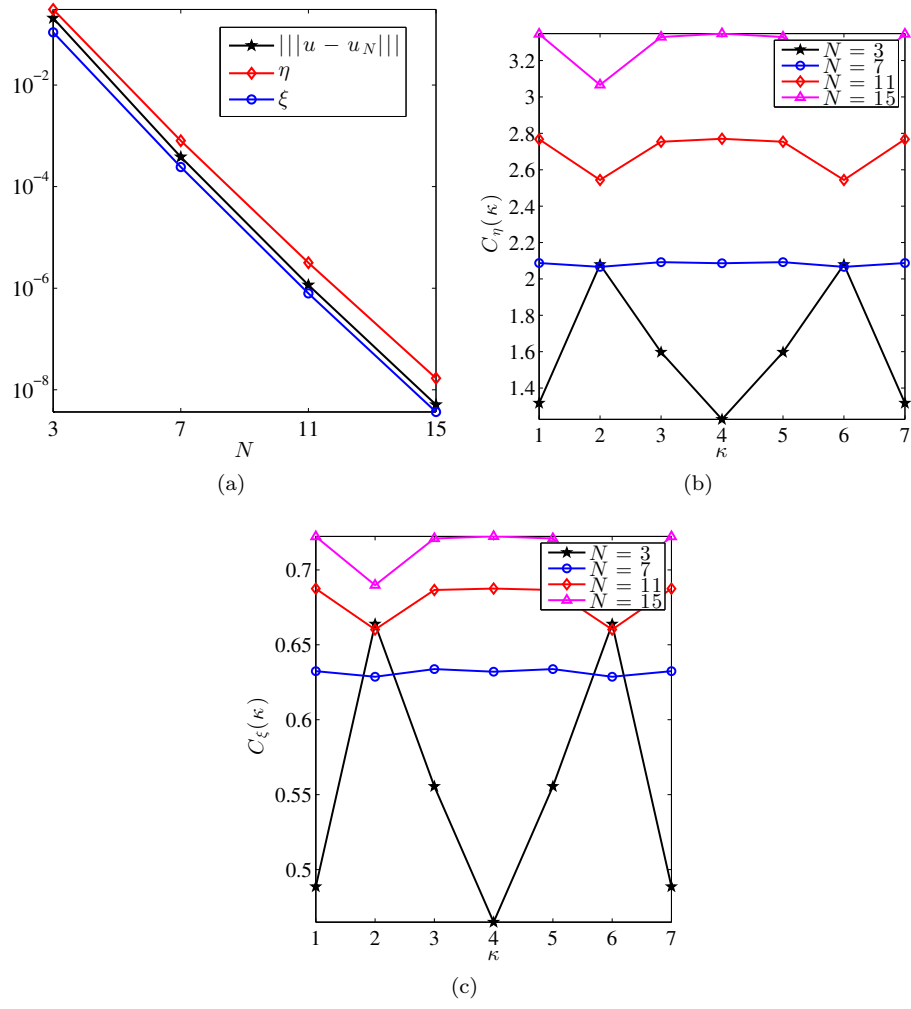
FIGURE 3. (a) Global error and the upper/lower bound estimator for $V(x) = 0.01$ and $f(x) = \sin(6x)$. (b) Local effectiveness of the upper bound characterized by $C_\eta(\kappa)$ for each element. (c) Local effectiveness of the lower bound characterized by $C_\xi(\kappa)$ for each element.
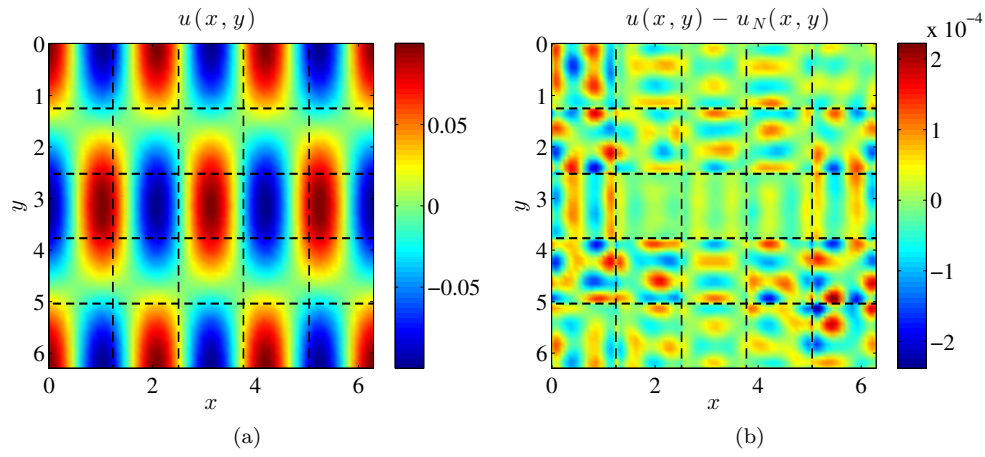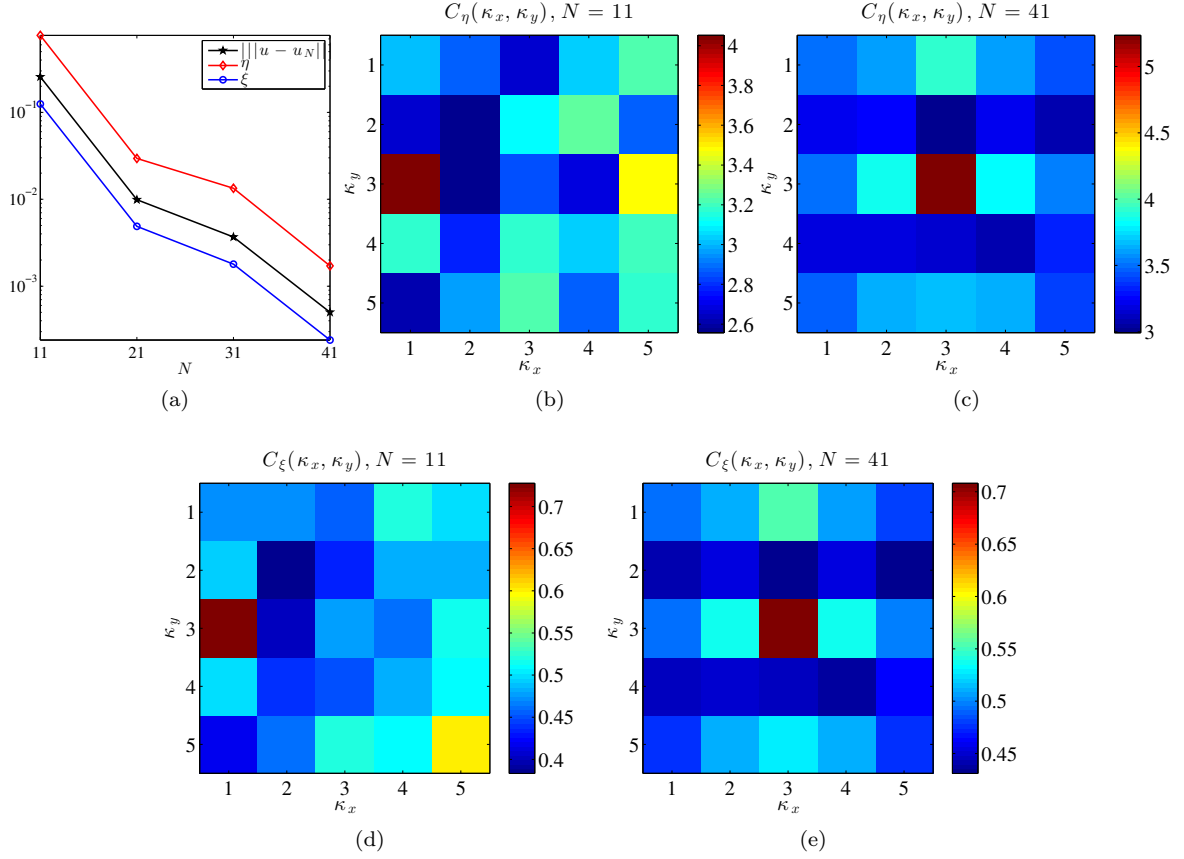
FIGURE 4. (a) The reference solution $u(x, y)$ corresponding to $V(x, y) = 0.01$ and $f(x, y) = \cos(3x) \cos(y)$. (b) Point-wise error between the reference solution $u(x, y)$ and the numerical solution $u_N(x, y)$ calculated using the ALB set with 5×5 elements and $N = 31$ basis functions per element.

FIGURE 5. (a) Global error and the upper/lower bound estimator for $V(x,y) = 0.01$ and $f(x,y) = \cos(3x)\cos(y)$. (b) Local effectiveness of the upper bound characterized by C_η in each element for $N = 11$. (c) Local effectiveness of the upper bound characterized by C_η in each element for $N = 41$. (d) Local effectiveness of the lower bound characterized by C_ξ in each element for $N = 11$. (e) Local effectiveness of the lower bound characterized by C_ξ in each element for $N = 41$.

6.3. **Indefinite operators**

We now demonstrate the effectiveness of the upper and lower bound estimates for indefinite operators. We start from a 1D example on a domain $\Omega = (0, 2\pi)$ with periodic boundary conditions. The potential function $V(x)$ is given by the sum of three Gaussians with negative magnitude, as shown in Fig. 6 (a). The operator $A = -\Delta + V$ has 3 negative eigenvalues and is indefinite. The right hand side is $f(x) = \sin(6x)$. The domain is partitioned into 7 elements for the ALB calculation. Fig. 6 (b) shows the reference solution u to Eq. (14), and Fig. 6 (c) shows the point-wise error $u - u_N$ using $N = 11$ ALBs per element.

(a)

(b)

(c)
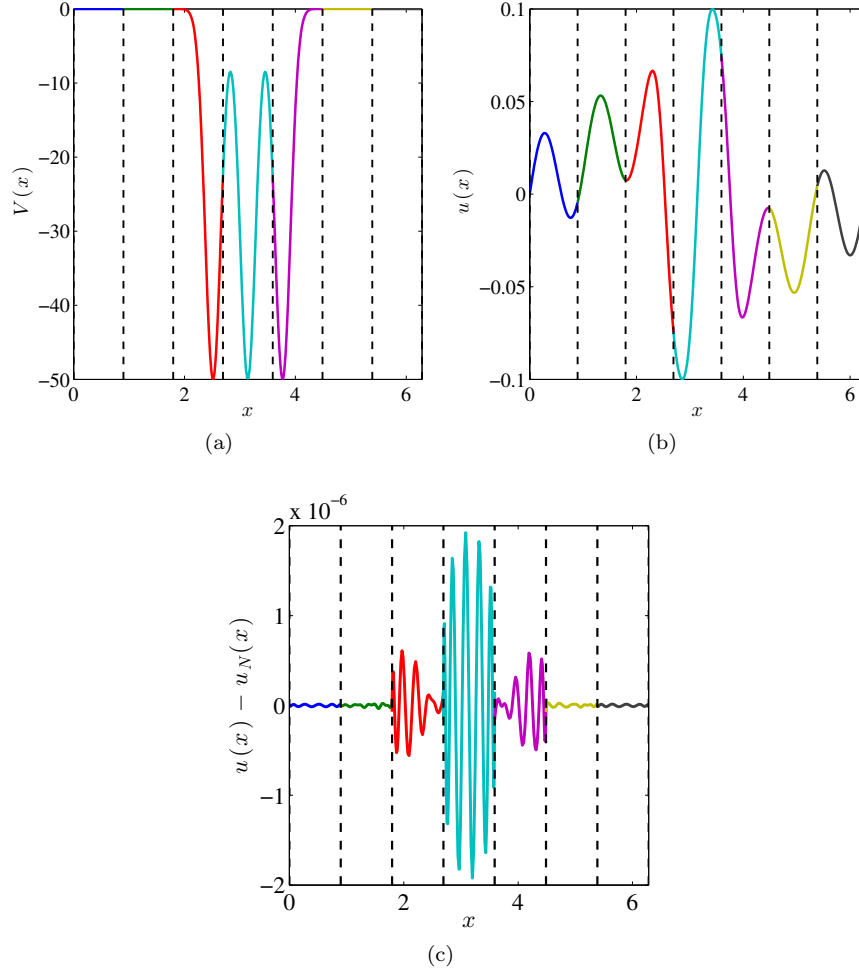
FIGURE 6. (a) The potential $V(x)$ given by the sum of three Gaussians with negative magnitude. (b) The reference solution $u(x)$ corresponding to the potential $V(x)$ in (a) and the right hand side $f(x) = \sin(6x)$. (c) Point-wise error between the reference solution $u(x)$ and the numerical solution $u_N(x)$ calculated using the ALB set with 7 elements and $N = 11$ basis functions per element.

Fig. 7 (a) shows the error in the energy norm, the computed upper and lower bound estimates as the number of ALBs per element N increases from 3 to 15. Similar to Fig. 3, the computed η and ξ are upper and lower bounds for the true error $\|u - u_N\|$ for all N across a wide range of accuracy. Furthermore, the computed ξ is always a lower bound of $\|u - u_N\|$ from $N = 3$ to $N = 15$. This is guaranteed by the property of the lower bound in Proposition 4.5.

We should note that when the number of basis functions is very small ($N = 3$), the accuracy is low and the ALB approximation is in its pre-asymptotic regime. In such case, the upper bound is very close

to the true error. In fact as indicated by Theorem 4.1, η may not even be a rigorous upper bound for highly indefinite operators with very few basis functions.
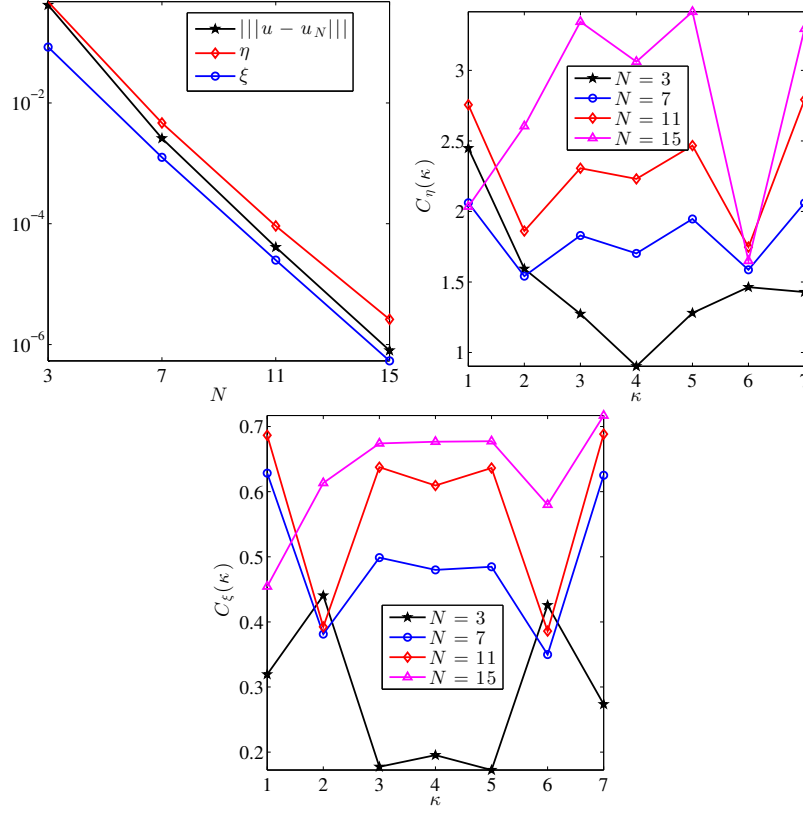

FIGURE 7. (a) Global error and the upper/lower bound estimator for $V(x)$ given in Fig. 6 (a) and $f(x) = \sin(6x)$. (b) Local effectiveness of the upper bound characterized by C_η in each element. (c) Local effectiveness of the lower bound characterized by C_η in each element.

Our final examples are two indefinite problems on a 2D domain $\Omega = (0, 2\pi) \times (0, 2\pi)$. The first problem is a homogeneous Helmholtz equation with $V(x, y) = -16.5$ and the operator $A = -\Delta + V$ has 49 negative eigenvalues. The right hand side is

$$f(x, y) = \exp(-2(x - \pi)^2 - 2(y - \pi)^2), \tag{58}$$

which is a Gaussian located at the center of Ω. The second problem is that V is given by the sum of four Gaussians with negative magnitude, as illustrated in Fig. 10 (a). The operator $A = -\Delta + V$ has 26 negative eigenvalues. The right hand side is chosen to be $f(x, y) = \cos(3x)\cos(y)$ satisfying the periodic boundary condition. For the first problem, Fig. 8 (b) shows the reference solution u to Eq. (14) and Fig. 8 (c) shows the point-wise error $u - u_N$ using $N = 31$ ALBs per element. In the ALB computation, the domain is partitioned into 5×5 elements, indicated by black dashed lines. Similarly for the second problem, Fig. 10 shows solution u to Eq. (14) and the point-wise error $u - u_N$ using $N = 31$ ALBs per element.

Fig. 9 (a)-(e) illustrates the global and local effectiveness of the upper and lower bound estimates for the Helmholtz problem, as the number of ALBs per element N increases from 21 to 51. Compared to the positive definite case in Fig. 5, the true error is larger using a comparable number of basis functions, reflecting that the Helmholtz equation is more difficult to solve. Nonetheless, η and ξ provide effective bounds for the true error in all cases. Similar results can be found for the indefinite example with negative Gaussian potentials in Fig. 11 (a)-(e). In all calculations, the computed lower bound estimator

remains a lower bound for the true error. In particular, the estimators still hold quite tightly in the pre-asymptotic regime ($N = 11$) where the ALB approximation is crude and has large numerical error.

(a)   (b)
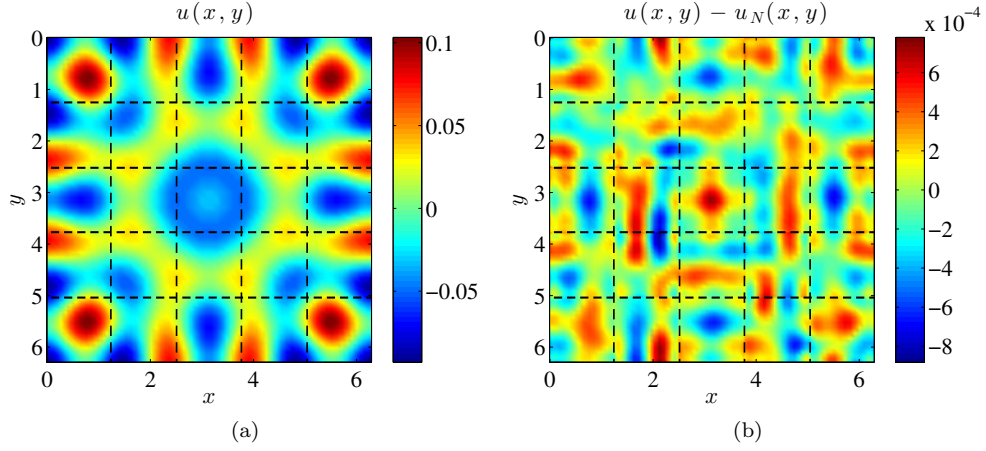
FIGURE 8. (a) The reference solution $u(x, y)$ corresponding to $V(x, y) = -16.5$ and $f(x, y)$ in Eq. (58), which is a Gaussian localized at the center of Ω. (b) Point-wise error between the reference solution $u(x, y)$ and the numerical solution $u_N(x, y)$ calculated using the ALB set with 5×5 elements and $N = 31$ basis functions per element.
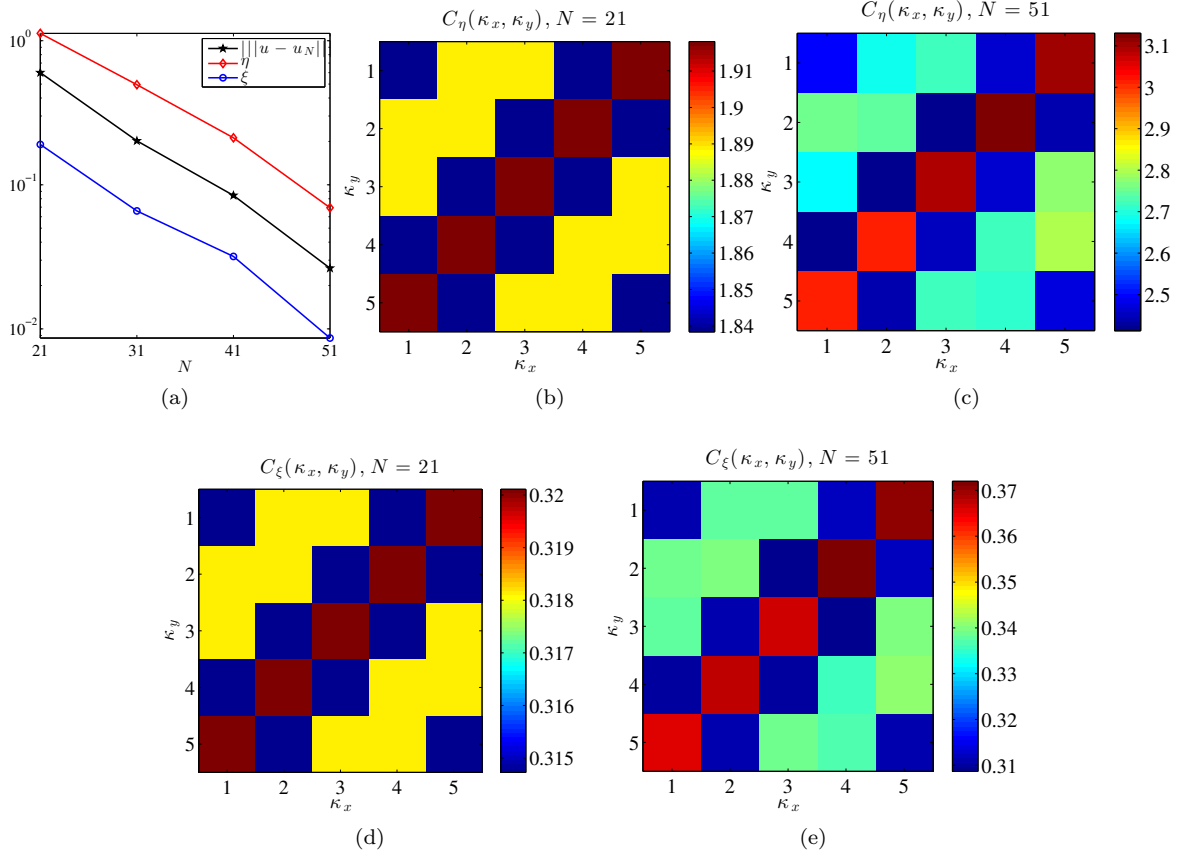
FIGURE 9. (a) Global error and the upper/lower bound estimator for $V(x,y) = -16.5$ and $f(x,y)$ in Eq. (58), which is a Gaussian localized at the center of Ω. (b) Local effectiveness of the upper bound characterized by C_η in each element for $N = 21$. (c) Local effectiveness of the upper bound characterized by C_η in each element for $N = 51$. (d) Local effectiveness of the lower bound characterized by C_ξ in each element for $N = 21$. (e) Local effectiveness of the lower bound characterized by C_ξ in each element for $N = 51$.
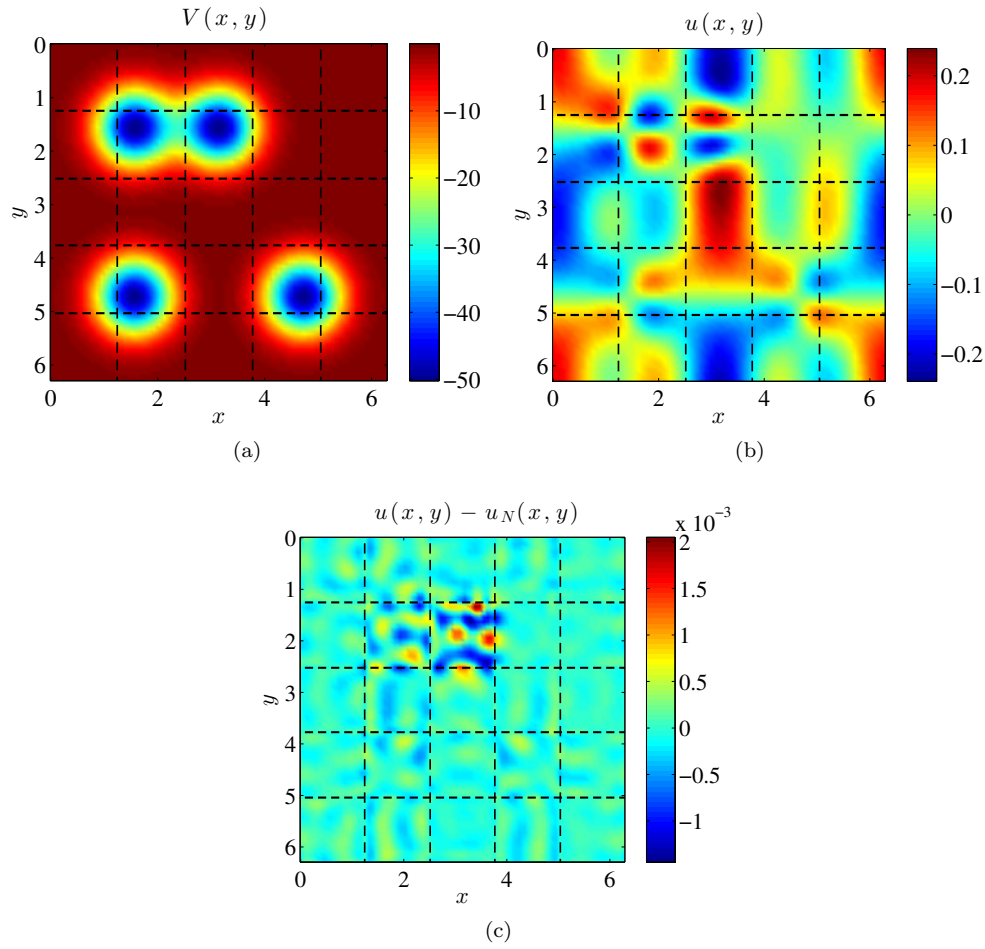
FIGURE 10. (a) The potential $V(x, y)$ four Gaussians with negative magnitude. (b) Solution $u(x, y)$ corresponding to $V(x, y)$ given in (a) and $f(x, y) = \cos(3x)\cos(y)$. (c) Point-wise error between the reference solution $u(x, y)$ and the numerical solution $u_N(x, y)$ calculated using the ALB set with 5×5 elements and $N = 31$ basis functions per element.
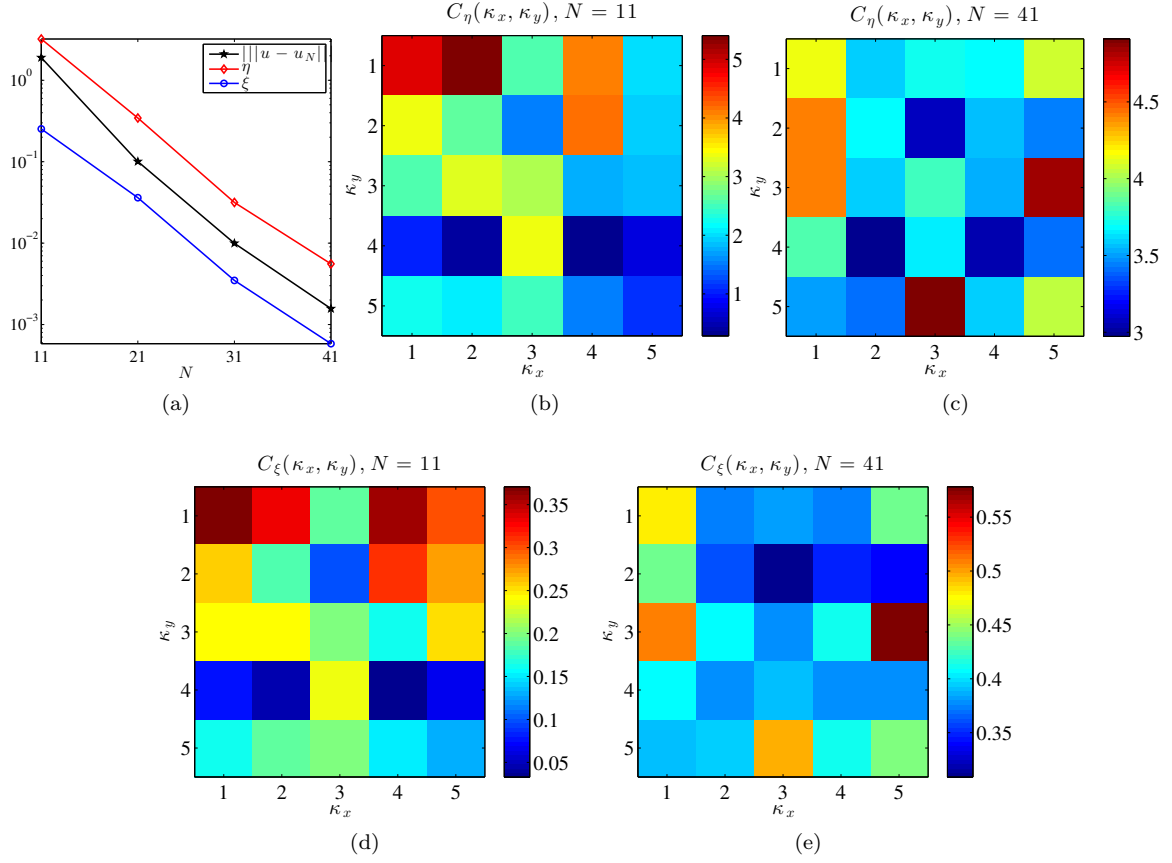
FIGURE 11. (a) Global error and the upper/lower bound estimator for $V(x, y)$ given in Fig. 10 (a) and $f(x, y) = \cos(3x)\cos(y)$. (b) Local effectiveness of the upper bound characterized by C_η in each element for $N = 11$. (c) Local effectiveness of the upper bound characterized by C_η in each element for $N = 41$. (d) Local effectiveness of the lower bound characterized by C_ξ in each element for $N = 11$. (e) Local effectiveness of the lower bound characterized by C_ξ in each element for $N = 41$.

6.4. **Justification of the treatment of $\mathrm{d}_\kappa^u(u_N)$**

In the numerical computation of the upper and lower bound estimates, we approximated the non-computable constant $\mathrm{d}_\kappa^u(u_N)$ by the computable constant d_κ. Below we provide numerical justification of such approximation by direct computation of $\mathrm{d}_\kappa^u(u_N)$ via the reference solution. We compare with d_κ and $\mathrm{b}_\kappa\gamma_\kappa$ since these three terms appear together in $\eta_{\mathrm{J},\kappa}$ in Eq. (12).

Fig. 12 (a) and (b) compare $\mathrm{d}_\kappa^u(u_N)$, d_κ and $\mathrm{b}_\kappa\gamma_\kappa$ for the positive definite and the indefinite 1D examples, respectively. We observe that the magnitude of $\mathrm{d}_\kappa^u(u_N)$ is comparable to that of d_κ. $\mathrm{b}_\kappa\gamma_\kappa$ is much smaller compared to $\mathrm{d}_\kappa^u(u_N)$ and d_κ. This is a direct consequence of Proposition 7.1, which states that b_κ is in general very small for 1D systems.
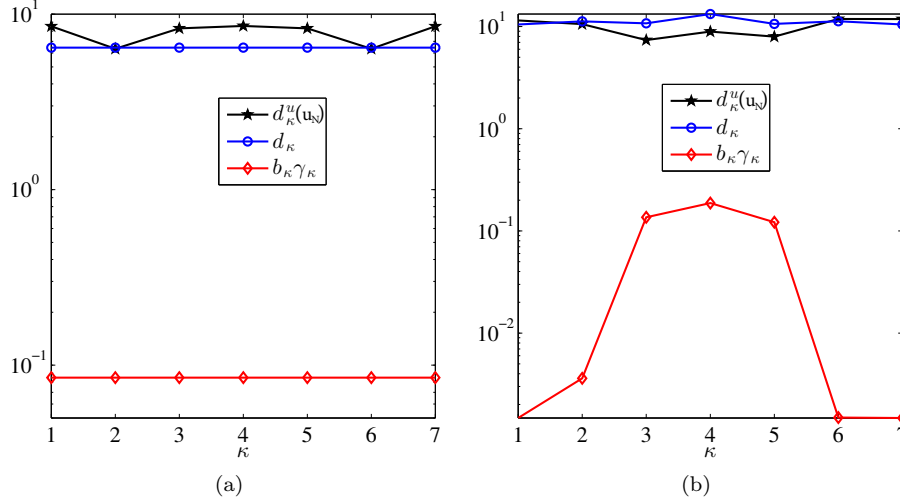

FIGURE 12. Comparison of $\mathrm{d}_\kappa^u(u_N)$, d_κ and $\mathrm{b}_\kappa\gamma_\kappa$ for (a) the positive definite case with $V(x) = 0.01$ with $N = 7$. (b) the indefinite case with $V(x)$ given in Fig. 6 (a) with $N = 7$.

Fig. 13 compare $\mathrm{d}_\kappa^u(u_N)$, d_κ and $\mathrm{b}_\kappa\gamma_\kappa$ for the positive definite case $V = 0.01$, the indefinite case $V = -16.5$, and the indefinite case with V given by the sum of negative Gaussians in Fig. 10 (a). In all cases, the magnitude of $\mathrm{d}_\kappa^u(u_N)$ is comparable to that of d_κ. Furthermore, both $\mathrm{d}_\kappa^u(u_N)$ and d_κ are much smaller compared to $\beta_\kappa\gamma_\kappa$. Therefore the effectiveness of the estimator remains unchanged even if $\mathrm{d}_\kappa^u(u_N)$ is neglected. We expect similar results can be observed for systems of higher dimensionality.

Finally we provide a second justification by comparing the total contribution of the jump term in the upper bound estimator

$$\eta_J^2 = \sum_\kappa \eta_{J,\kappa}^2,$$

and the total contribution of the jump term in the energy norm

$$E_J = \sum_\kappa \tfrac{\gamma_\kappa}{2} \|\llbracket u_N \rrbracket\|_{\partial\kappa}^2.$$

This is given in Table 1. It shows that the approximation $\mathrm{d}_\kappa^u(u_N) \approx \mathrm{d}_\kappa$ does not lead to underestimation of the jump term, which is consistent with the observation in Fig. 12 and 13.
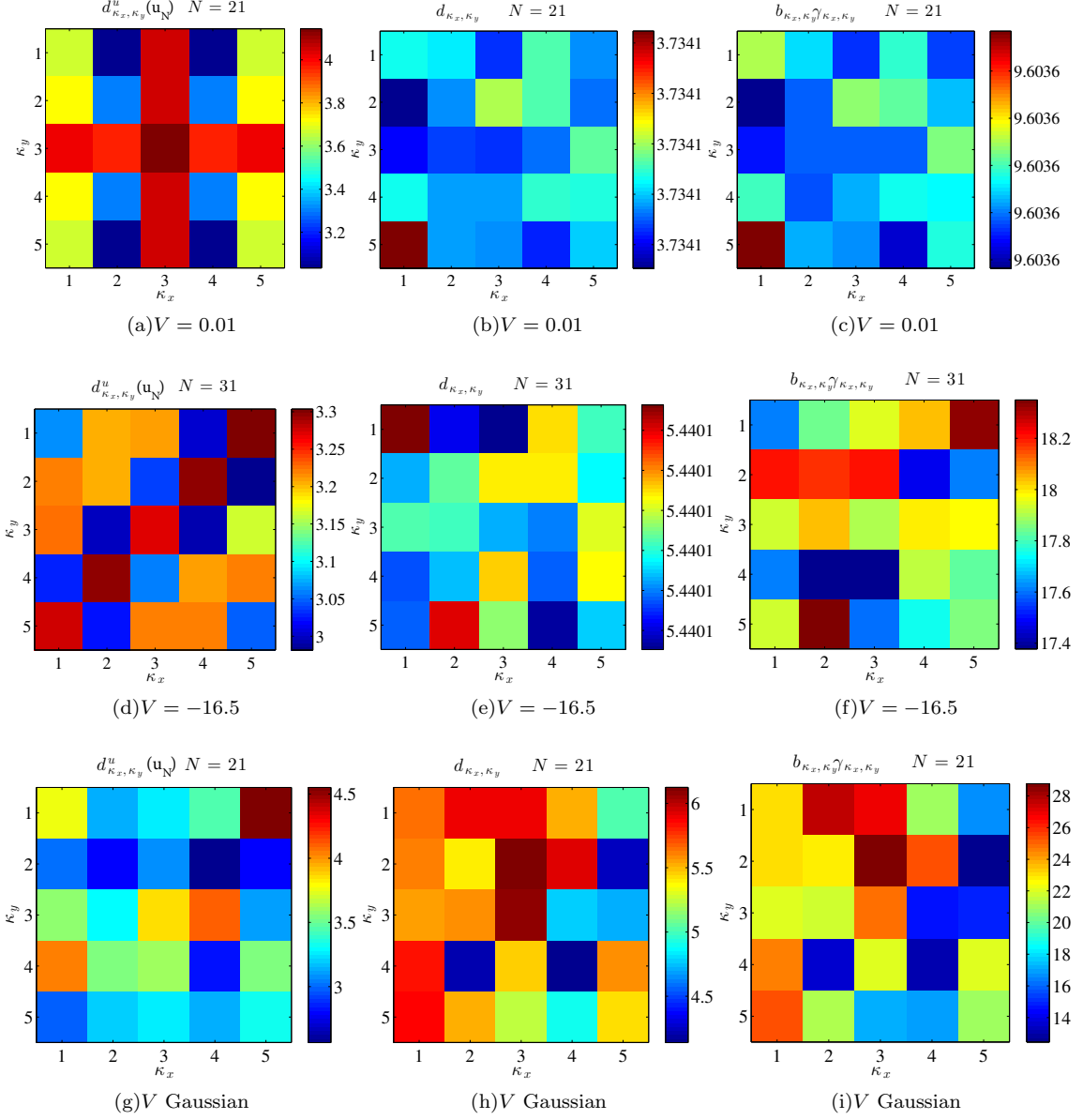
FIGURE 13. Comparison of $\mathtt{d}_\kappa^u(u_N)$, $\mathtt{d}_\kappa$ and $\mathtt{b}_\kappa\gamma_\kappa$ for 2D test problems for (a-c) the positive definite case $V = 0.01$ (d-f) the indefinite case $V = -16.5$ (g-i) the indefinite case with V given by the sum of negative Gaussians in Fig. 10 (a).

Problem	N	E_J	η_J^2
1D $V = 0.01$	7	2.0179×10^{-8}	2.0182×10^{-8}
2D $V = 0.01$	21	1.2030×10^{-5}	9.1593×10^{-5}
1D Gaussian	11	6.4687×10^{-11}	6.4697×10^{-11}
2D $V = -16.5$	31	4.7352×10^{-3}	5.6649×10^{-2}
2D Gaussian	21	1.6226×10^{-3}	2.8348×10^{-2}

TABLE 1. Comparison of the total contribution of the jump term in the estimator η_J^2, and the total contribution of the jump term in the energy error E_J.

7. Conclusion and future work

We present the first systematic work for deriving a posteriori error estimates for general non-polynomial basis functions in a interior penalty discontinuous Galerkin (DG) formulation for solving second order linear PDEs. The estimates not only serve to quantify the error sharply for a given computation, but also can lead an adaptive algorithm to refine the elements non-uniformly by adding (or even removing/coarsening) basis functions to certain elements. This allows a best approximation for a given number of degrees of freedom in order to reduce the computing time even when relatively few degrees of freedom are employed. A non-uniform distribution of the number of local basis functions is in this case mandatory to develop powerful solvers, in particular when inhomogeneous data of the PDE is involved. It turns out that the standard polynomial hp DG-method may benefit from this analysis as it involves numerically computed constants.

Our analysis requires the exact solution to lie in $H^2(\kappa)$ for each element κ which may seem limiting when dealing with a posteriori estimates for Poisson's equation as a uniform refinement leads to optimal convergence rates in the asymptotic limit. We remark that despite the above asymptotic reasoning there are numerous cases where the a posteriori analysis for regular functions is still interesting, for example if the PDE involves a strong small-scale character (but still being smooth) either due to strongly oscillating material coefficients or a wave-like character of the underlying PDE (Helmholtz equation for instance). Or, if the data of the PDE and thus the solution as well has an inhomogeneous character so that a uniform refinement involves too many degrees of freedom. In this case, combining the estimates with an adaptive algorithm as outlined above will result in an optimal balance of degrees of freedom per element.

Our framework for developing explicitly computable constants for a posteriori error estimates are not limited to second order PDEs, nor it is necessarily limited to discontinuous Galerkin framework. In a forthcoming publication we will demonstrate the method for eigenvalue problems. It is also possible to generalize the method to multiscale methods and reduced basis methods.

Acknowledgments

Appendix

Proposition 7.1. *Let $\kappa = [a, b]$ be a 1D element and $\mathbb{V}_N(p; \kappa) = \text{span}\{x^j, j \leq p\}$ be the function space spanned by polynomials with degree less than or equal to p. Then $\forall p \geq 2$, $\mathtt{b}_\kappa = 0$.*

Proof. Define $c = (a + b)/2$. For any $v \in H^1(\kappa)$, $v \perp \mathbb{V}_N(p; \kappa)$ with $p \geq 2$, we have

$$(v, 1)_{\star, \kappa} = 0, \quad (v, (x - c))_{\star, \kappa} = 0, \quad (v, (x - c)^2)_{\star, \kappa} = 0.$$

Using the definition of the inner product $(\cdot, \cdot)_{\star, \kappa}$

$$\int_a^b v(x)\, dx = 0, \quad \int_a^b v'(x)\, dx = 0, \quad \int_a^b v'(x)(x - c)d\, x = 0.$$

With integration by parts, we have $v(a) = v(b) = 0$. Therefore $\|v\|_{\partial \kappa} = 0$. Using the definition of $\mathtt{b}_\kappa$ we obtain $\mathtt{b}_\kappa = 0$. $\qquad\square$

References

[1] M. Ainsworth and R. Rankin, *Technical note: A note on the selection of the penalty parameter for discontinuous Galerkin finite element schemes*, Numer. Methods Partial Differential Equations, 28 (2012), pp. 1099–1104.

[2] M. Amara, R. Djellouli, and C. Farhat, *Convergence analysis of a discontinuous Galerkin method with plane waves and Lagrange multipliers for the solution of Helmholtz problems*, SIAM J. Numer. Anal., 47 (2009), pp. 1038–1066.

[3] D. N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742 – 760.

[4] I. Babuška and M. Zlámal, *Nonconforming elements in the finite element method with penalty*, SIAM J. Numer. Anal., 10 (1973), pp. 863 – 875.

[5] G. A. Baker, *Finite element methods for elliptic equations using nonconforming elements*, Math. Comp., 31 (1977), pp. 45–59.

[6] C. E. Baumann and J. T. Oden, *A discontinuous hp finite element method for convection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 175 (1999), p. 311–341.

[7] W. E and B. Engquist, *The heterognous multiscale methods*, Comm. Math. Sci., 1 (2003), pp. 87–132.

[8] J. Douglas, Jr. and T. Dupont, *Interior penalty procedures for elliptic and parabolic Galerkin methods*, Computing methods in applied sciences (Second Internat. Sympos., Versailles, 1975), 58 (1976), pp. 207–216.

[9] Y. Epshteyn and B. Rivière, *Estimation of penalty parameters for symmetric interior penalty Galerkin methods*, J. Comput. Appl. Math., 206 (2007), pp. 843–872.

[10] M. J. Frisch, J. A. Pople, and J. S. Binkley, *Self-consistent molecular orbital methods 25. supplementary functions for gaussian basis sets*, J. Chem. Phys., 80 (1984), pp. 3265–3269.

[11] S. Giani and E. J. C. Hall, *An a posteriori error estimator for hp-adaptive discontinuous Galerkin methods for elliptic eigenvalue problems*, Math. Mod. Meth. Appl. Sci., 22 (2012), pp. 1250030–1250064.

[12] R. Hiptmair, A. Moiola, and I. Perugia, *Plane wave discontinuous Galerkin methods for the 2D Helmholtz equation: analysis of the p-version*, SIAM J. Numer. Anal., 49 (2011), pp. 264–284.

[13] T. Y. Hou and X.-H. Wu, *A multiscale finite element method for elliptic problems in composite materials and porous media*, J. Comput. Phys., 134 (1997), pp. 169–189.

[14] P. Houston, D. Schötzau, and T. P. Wihler, *Energy norm a posteriori error estimation of hp-adaptive discontinuous Galerkin methods for elliptic problems*, Math. Mod. Meth. Appl. Sci., 17 (2007), pp. 33–62.

[15] P. Houston, C. Schwab, and E. Süli, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 39 (2002), pp. 2133–2163.

[16] J. Junquera, O. Paz, D. Sanchez-Portal, and E. Artacho, *Numerical atomic orbitals for linear-scaling calculations*, Phys. Rev. B, 64 (2001), pp. 235111–235119.

[17] O. A. Karakashian and F. Pascal, *A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 2374–2399.

[18] J. Kaye, L. Lin, and C. Yang, *A posteriori error estimator for adaptive local basis functions to solve Kohn-Sham density functional theory*, Commun. Math. Sci, in press, (2015).

[19] A. V. Knyazev, *Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method*, SIAM J. Sci. Comp., 23 (2001), p. 517.

[20] L. Lin, J. Lu, L. Ying, and W. E, *Adaptive local basis set for Kohn-Sham density functional theory in a discontinuous Galerkin framework I: Total energy calculation*, J. Comput. Phys., 231 (2012), pp. 2140–2154.

[21] J. Nitsche, *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*, Abh. Math. Sem. Univ. Hamburg, 36 (1971), pp. 9–15.

[22] M. Ohlberger, *A posteriori error estimates for the heterogeneous multiscale finite element method for elliptic homogenization problems*, Multiscale Model. Simul., 4 (2005), pp. 88–114.

[23] D. Schötzau and L. Zhu, *A robust a-posteriori error estimator for discontinuous Galerkin methods for convection–diffusion equations*, Appl. Numer. Math., 59 (2009), pp. 2236–2255.

[24] C. Schwab, *p-and hp-Finite Element Methods*, Oxford Univ. Pr., New York, 1998.

[25] B. Stamm and T. Wihler, *hp-Optimal discontinuous Galerkin methods for linear elliptic problems*, Math. Comp., 79 (2010), pp. 2117–2133.

[26] R. Tezaur and C. Farhat, *Three-dimensional discontinuous Galerkin elements with plane waves and Lagrange multipliers for the solution of mid-frequency Helmholtz problems*, Int. J. Numer. Meth. Eng., 66 (2006), pp. 796–815.

[27] M. F. Wheeler, *An elliptic collocation-finite element method with interior penalties.*, SIAM J. Numer. Anal., 15 (1978), p. 152–161.